TECHNICKÁ UNIVERZITA V LIBERCI

Fakulta mechatroniky, informatiky a mezioborových studií

Studijní program: N2612 – Elektrotechnika a informatika

Studijní obor: 1802T007 – Informační technologie

Vliv řečníka a přenosového kanálu na systém rozpoznávání řeči

Speaker and transmission channel effect on speech recognition system

Diplomová práce

Autor: **Bc. Michaela Kuchařová**
Vedoucí práce: **Prof. Ing. Jan Nouza, CSc.**
Konzultant: **Ing. Petr Červa, Ph.D.**

V Liberci 1. 5. 2011

Originální zadání

Prohlášení

Byla jsem seznámena s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 o právu autorském, zejména § 60 (školní dílo).

Beru na vědomí, že TUL má právo na uzavření licenční smlouvy o užití mé diplomové práce a prohlašuji, že **s o u h l a s í m** s případným užitím mé diplomové práce (prodej, zapůjčení apod.).

Jsem si vědoma toho, že užít své diplomové práce či poskytnout licenci k jejímu využití mohu jen se souhlasem TUL, která má právo ode mne požadovat přiměřený příspěvek na úhradu nákladů, vynaložených univerzitou na vytvoření díla (až do jejich skutečné výše).

Diplomovou práci jsem vypracovala samostatně s použitím uvedené literatury a na základě konzultací s vedoucím diplomové práce a konzultantem.

Datum

Podpis

Poděkování

Ráda bych poděkovala panu Prof. Ing. Janu Nouzovi, CSc. za jeho rady, ochotu a čas strávený konzultacemi mojí diplomové práce. Také bych mu chtěla poděkovat za to, že mi umožnil přístup k profesionálnímu programu pro rozpoznávání mluvené češtiny a nejnovějším modelům, které mi umožnily získat zajímavé výsledky mé práce. Také bych zde ráda poděkovala panu Ing. Petru Červovi, Ph.D. za jeho pomoc a ochotu při konzultacích ohledně metod adaptace.

Abstrakt

Tato diplomová práce se zabývá závislostí úspěšnosti rozpoznávání mluvené řeči na použitém mikrofonu a částečné též na mluvčím. Cílem práce bylo vytvořit databázi, která by obsahovala nahrávky od různých mluvčích a z různých mikrofonů, které by byly vhodné pro objektivní porovnání daných mikrofonů. Po zjištění, že rozpoznávání řeči je poměrně značně závislé na použitém mikrofonu, jsem vytvořila systém pro rozpoznávání řeči v prostředí HTK (The Hidden Markov Model Toolkit) a otestovala jsem základní metody a různé vstupní parametry tohoto systému.

Jako první metoda adaptace byla otestována intuitivní metoda, která přidávala k trénovací sadě nahrávek adaptační data s různou vahou. Poté jsem vyzkoušela několik pokročilejších metod adaptace a ověřovala, jak se po jejich aplikaci změní rozdíl úspěšností mezi nahrávkami z různých mikrofonů. Toto testování proběhlo jak na rozpoznávacím systému v prostředí HTK, tak i na profesionálním systému pro rozpoznávání řeči, který se používá v praxi v několika komerčních aplikacích a byl vyvinut Laboratoří počítačového zpracování řeči na Ústavu informačních technologií a elektroniky, FM TUL.

Nejlepší testovaná metoda MLLR (Maximum Likelihood Linear Regression) dosáhla s rozpoznávacím systémem poskytnutým Laboratoří počítačového zpracování řeči průměrné zlepšení úspěšnosti rozpoznávání řeči 2 %. Vzhledem k relativně malému objemu adaptačních dat (jednalo se o v průměru 30 vteřin dlouhou nahrávku, která obsahovala foneticky bohaté věty) je uvedené zlepšení znatelné.

Klíčová slova: zpracování řeči, rozpoznávání řeči, druhy mikrofonů, adaptace na mluvčího, adaptace na přenosový kanál

Abstract

This master thesis deals with the dependence of success in speech recognition on a used microphone and partly on a speaker. The aim was to create a database that would contain recordings from different speakers and from different microphones, which would be suitable for objective comparison of the microphones. After finding that speech is significantly dependent on the microphone, I have created a system for speech recognition in the HTK (The Hidden Markov Model Toolkit) and I tested the basic methods and different input parameters of the system.

As a first adaptation method, I tested an intuitive method, which added adaptation data (with different weights) to the training set. Then I tested a few more advanced methods of adaptation and investigated how the difference changed between records from different microphones after their application. Testing was done with a recognition system in the HTK, and on the professional system for speech recognition, which is used in several commercial applications and which was developed by the Laboratory of speech processing at the Institute of Information Technology and Electronics at Technical University Liberec.

The best tested method was MLLR (Maximum Likelihood Linear Regression). It achieved an average improvement of speech recognition accuracy in range of 2 %. As the amount of the adaptation data was rather small (it was in average 30 seconds long record, which contained phonetically rich sentences), the improvement is good.

Keywords: speech processing, speech recognition, types of microphones, speaker adaptation, transmission channel adaptation

Obsah

Slovník zkratek

HTK …… The Hidden Markov Model Toolkit
(sestava programů pro vývojovou a experimentální práci s markovskými modely)

HMM …… Hidden Markov Model
(skrytý markovský model)

SI ……Speaker Independent
(nezávislý na mluvčím)

SD …… Speaker Dependent
(závislý na mluvčím)

SA …… Speaker Adapted
(adaptovaný na mluvčího)

MAP ……Maximum A Posteriori
(maximální aposteriorní pravděpodobnost)

MLLR …… Maximum Likelihood Linear Regression
(maximálně věrohodná lineární regrese)

CMLLR …… Constrained maximum likelihood linear regression
(omezená maximálně věrohodná lineární regrese)

SAT …… Speaker Adaptive Training
(trénování pro účely adaptace na mluvčího)

MFCC …… Mel-frequency cepstral coefficients
(melovské frekvenční kepstrální koeficienty)

1. Úvod

Počítačové rozpoznávání řeči představuje moderní vědní disciplínu zaměřenou na podporu hlasové komunikace a interakce s počítačem. Rozvoj této disciplíny byl zahájen v 60. letech, v době vzniku prvních výpočetních systémů. V 70. a 80. letech byly položeny základy většiny metod, na nichž jsou současné systémy rozpoznávání řeči založeny. Od 90. let se objevují první komerční systémy. Nejprve to byly jednoduché programy pro hlasové ovládání počítače, následovaly programy umožňující nejdříve izolované a později spojité diktování do počítače, načež se pozornost vědců i firem zaměřila na nejsložitější úlohy, k nimž patří např. titulkování televizních pořadů, přepis jednání či zpracování zvukových archivů.

Rozpoznávání řeči je závislé na konkrétním jazyku. První systémy byly vytvářeny pro velké světové jazyky, jako jsou angličtina, němčina, španělština či japonština. Bylo to dáno dvěma faktory. Zaprvé velikostí trhu, na němž bylo možné hotové programy uplatnit a zaplatit tak nemalé náklady na vývoj. Druhým faktorem byla složitost jazyka. Všechny výše uvedené jazyky, a platí to zejména pro angličtinu, se vyznačují poměrně malou mírou ohebnosti (inflexe). Většina slov existuje buď v jediném základním tvaru, nebo v malém počtu tvarů odvozených podle relativně jednoduchých pravidel (např. přidáním koncovky –s v případě množného čísla podstatných jmen v angličtině). Pro rozpoznávání řeči to znamená velkou výhodu, protože slovník systému si vystačí s řádově desítkami tisíc položek.

Čeština patří naopak k jazykům, které mají velmi bohatou morfologii. Podstatná jména, přídavná jména, zájmena a číslovky se skloňují, slovesa časují, a počet různých slovních tvarů se pohybuje ne v desítkách, ale spíše ve stovkách tisíc. To významnou měrou komplikuje vývoj systémů rozpoznávání řeči pro ohebné jazyky. První systémy určené pro češtinu byly vytvořeny v polovině 90. let [Hájek94]. V té době se jednalo o poměrně jednoduché programy pro ovládání počítače, určené primárně pro osoby s motorickým postižením. Přibližně o 10 let později vznikly první české diktovací programy [Nouza00] a později i další aplikace, např. programy pro přepis televizních pořadů [Nouza05] nebo hlasový vstup pro mobilní telefony.

Výzkum a vývoj však dále pokračuje. Jedním z nejdůležitějších cílů je dosáhnout vyšší úspěšnosti rozpoznávání. Významnou roli zde hraje robustnost použitých metod, tj. schopnost pracovat spolehlivě za různých podmínek nasazení. Je známo, že přesnost rozpoznávání úzce souvisí s kvalitou signálu a ta je zase ovlivňována použitým mikrofonem,

zvukovou kartou a samozřejmě také prostředím, v němž je řeč zaznamenávána. Ještě větší roli ale hraje samotný mluvčí: jeho výslovnost, rychlost mluvy, volba slovníku a v neposlední řadě také případné vady řeči. Tato práce se zaměřuje primárně na otázky spojené s vlivem mikrofonu (a přenosového kanálu) a sekundárně též s vlivem mluvčího na úspěšnost v různých úlohách rozpoznávání řeči a na možnosti eliminace těchto jevů.

Zadání diplomové práce bylo motivováno výzkumnými projekty řešenými v Laboratoři počítačového zpracování řeči na Technické univerzitě v Liberci. Jsou zde vyvíjeny programy pomáhající hendikepovaným osobám, dále programy pro diktování do počítače (pro češtinu, slovenštinu a některé další jazyky), systémy pro přepis televizních a rozhlasových pořadů, či komplexní platformy pro přepis jednání či vysokoškolských přednášek. Všechny tyto aplikace vyžadují vysokou míru robustnosti. U aplikací určených pro osoby s tělesným postižením jde zejména o to vypořádat se s nestandardní výslovností některých uživatelů, u diktovacích programů je třeba minimalizovat vliv konkrétního mikrofonu a konkrétního hlasu na rozpoznaný text a u přepisovacích aplikací je nutné zvládnout časté změny mluvčích v rámci jediné nahrávky.

2. Cíle práce

Hlavním cílem práce je zjistit a kvantifikovat vliv různých mikrofonů na úspěšnost metod rozpoznávání řeči, a to zejména těch, které se používají v diktovacích systémech vyvíjených v Laboratoři počítačového zpracování řeči na Technické univerzitě v Liberci. Pro tento účel bylo pořízeno několik mikrofonů, které v této aplikaci přicházejí v úvahu. Jsou to v převážné míře mikrofony spojené s náhlavními sluchátky (tzv. head set) s připojením na USB port. Jedná se většinou o mikrofony střední třídy v cenové hladině od několika set do jednoho tisíce Kč. Úkoly stanovené zadáním a po dohodě s vedoucím práce jsou následující:

1) Vytvořit databázi nahrávek od různých mluvčích pomocí různých mikrofonů. Specifickou vlastností těchto nahrávek má být to, že řeč každé osoby bude paralelně nahrávána pomocí dvou mikrofonů, referenčního a testovaného. Tímto způsobem bude možné objektivně vyhodnotit rozdíl v úspěšnosti rozpoznávání mezi referenčním a každým z testovaných mikrofonů.
2) Výše uvedené nahrávky připravit tak, aby pro každou osobu a každý mikrofon existovala vždy část nahrávek určených pro adaptaci systému a druhá (větší) část použitelná pro testovací účely.
3) S využitím programů vyvinutých v Laboratoři počítačového zpracování řeči vyhodnotit vliv různých mikrofonů na úspěšnost rozpoznávání řeči, a to při různých akustických modelech (zejména pro tzv. monofony a trifony).
4) Seznámit se s prostředím HTK a vytvořit v něm několik programových modulů (pro trénování akustických modelů a pro rozpoznávání izolované i spojité řeči), na nichž se vyzkouší a ověří základní metody a parametry pro rozpoznávání mluvené češtiny, a to včetně jednoduché adaptace s využitím malého množství dat.
5) Seznámit se s metodami MAP a MLLR pro obecnou adaptaci akustického kanálu a v prostředí HTK je experimentálně otestovat na pořízené databázi nahrávek
6) Vybrané metody s nejlepšími parametry použít na skutečná data v rámci skutečného rozpoznávacího systému vyvíjeného v Laboratoři počítačového zpracování řeči.
7) Podrobně zdokumentovat všechny experimenty a jejich výsledky a navrhnout, jak na jejich základě zvýšit robustnost rozpoznávání řeči, zejména u diktovacích programů.

3. Mluvená řeč a principy jejího automatického rozpoznávání

Mluvená forma jazyka neboli řeč, je pro člověka nejpřirozenějším způsobem komunikace. Můžeme s její pomocí rychle vyjádřit jakoukoliv myšlenku. Druhá podoba jazyka, která je mnohem mladší, je textová forma. Psaná podoba se řídí pravidly daného jazyka, kdežto mluvená je mnohem různorodější (viz například rozdíly mezi spisovnou a nespisovnou formou jazyka). V počítačovém rozpoznávání řeči nás zajímá mluvená forma, která je vzhledem k různorodosti řeči obtížnější.

Nejmenším prvkem řeči je hláska (neboli foném), v textové podobě je nejmenší jednotkou písmeno (grafém). Zápis jednoho slova pomocí grafémů a fonémů se může velmi lišit, záleží na konkrétním jazyku. Hláska sama o sobě nenese žádný význam, ale skupina hlásek tvořící slovo, je už základní významovou jednotkou jazyka. Pokud se ve slově změní jedna hláska, může dojít ke změně významu slova (les x pes). Ještě ucelenějším objektem, který nese konkrétní informaci, je věta. Člověk potřebuje znát kontext, aby pochopil myšlenku. Slovo „pět“ sice nese informaci, ale samo o sobě (bez kontextu) člověku nic neříká. Kontext obsahuje právě věta. Pokud slyšíme větu „Je pět hodin“, dokážeme s ní již pracovat tak, jak je potřeba.

3.1 Mluvená čeština

Český jazyk patří mezi jazyky s volným slovosledem. To znamená, že pořadí jednotlivých slov ve větě není pevně dáno, ale je závislé na kontextu. Pevný slovosled mají například germánské jazyky (včetně angličtiny), u nich je dáno poměrně striktní řazení jednotlivých větných členů. Na rozdíl od jazyků s pevným slovosledem jsou jazyky s volným slovosledem obvykle flektivní, neboli ohebné. Čeština patří mezi flektivní jazyky a vyznačuje se bohatstvím slovních tvarů. Rozlišuje 10 slovních druhů, z nichž pět je ohebných a čtyři z nich se skloňují podle sedmi různých pádů. Vzhledem k těmto vlastnostem se počet slov používaných v češtině pohybuje v řádu stovek tisíc.

Čeština obsahuje 40 fonémů, tabulku všech fonémů je možné nalézt v [Nouza97]. Jednotlivé fonémy se zapisují pomocí znaků české abecedy. Fonetický přepis je pro češtinu ve

většině případů snadný, je možné aplikovat několik základních pravidel pro přepis (například: slovo „mě“ se foneticky přepíše na „mňe“). Především u slov převzatých z cizích jazyků je fonetický přepis obtížnější. Automatické programy vytvářející fonetický přepis se nazývají G2P (grapheme-to-phoneme) a řídí se kontextovými pravidly (kolem 30 základních pravidel aplikovatelných na slova českého původu). Vytvoření programu, který vrací přesný fonetický přepis je velmi složitá záležitost, více se o tomto tématu dá nalézt například v [Nouza09]. Na rozdíl od češtiny jsou jazyky jako angličtina či francouzština velmi složité pro fonetický přepis.

Jedno slovo je možné reprezentovat více výslovnostními variantami (to znamená, že existuje více možností fonetického přepisu). Například slovo „sedm“ je možné foneticky přepsat jako „sedm“ ale i „sedum“.

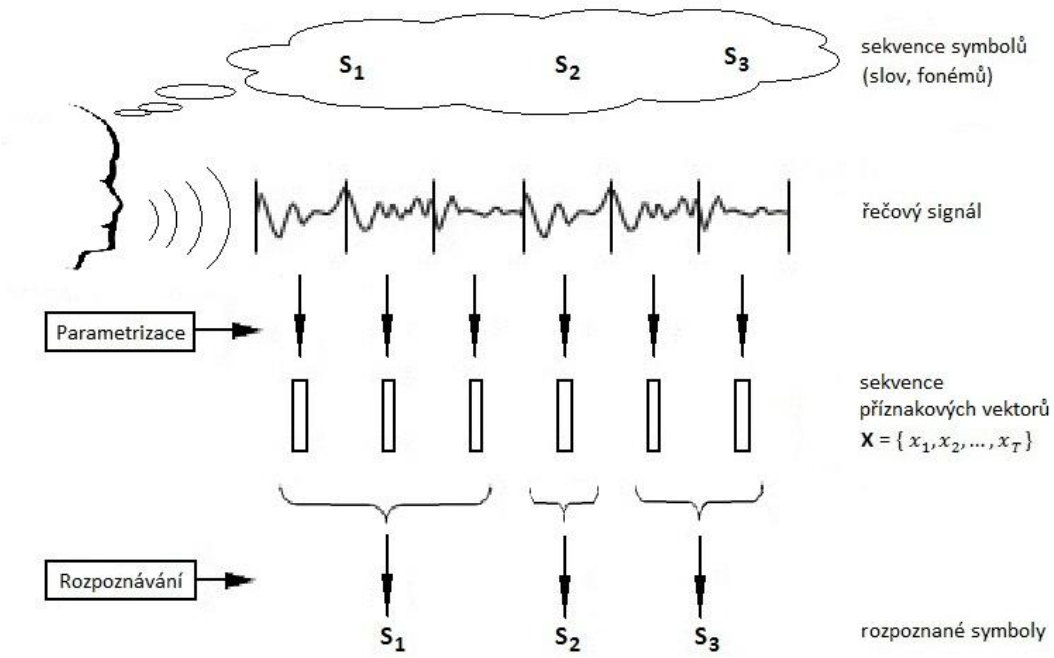
Reálné nahrávky ovšem neobsahují pouze fonémy daného jazyka, ale také různé mimoslovní ruchy. Mezi ně patří jak ruchy vyvolané řečovým ústrojím člověka, jako například nádech a výdech, mlasknutí a podobně, ale také ruchy pocházející z okolního prostředí jako je například hluk projíždějícího auta, muzika v pozadí, cvaknutí dveří apod. Je potřeba, aby byl systém na tyto mimoslovní ruchy připraven, a většina moderních systémů pro rozpoznávání řeči obsahuje modely nejenom pro fonémy, ale také pro nejvýznamnější a nejčastější ruchy, aby je bylo možné v dané promluvě detekovat.

3.2 Základy automatického rozpoznávání řeči

V automatickém rozpoznávání řeči se zabýváme dvěma hlavními úlohami. První úlohou je automatické rozpoznávání izolovaných slov. Tato úloha počítá s tím, že v dané promluvě, která se má rozpoznat a přiřadit k některému slovníkovému výrazu je pouze jedna slovníková položka (jedno slovo, popřípadě fráze – například „Jablonec nad Nisou“). Druhou, podstatně obtížnější úlohou, je rozpoznávání spojité řeči. Rozpoznávání spojité řeči je obtížnější hned z několika důvodů. U izolovaných slov předem víme, že v promluvě rozpoznáváme právě jednu slovníkovou položku, která bude začínat na začátku promluvy (popřípadě po oddělovači – obvykle jím bývá ticho) a po jejím skončení bude následovat konec nahrávky (popřípadě oddělovač). Zatímco u spojité řeči nejsou mezi jednotlivými slovy oddělovače. U spojité řeči navazuje jedno slovo za druhým a my jednak nevíme, které slovo bylo řečeno, ale ani nevíme, ve kterém okamžiku dané slovo začíná.

Izolovaná slova se mluvčí snaží pronášet srozumitelně, dává si větší pozor na výslovnost než u slov ve spojité řeči, kdy se často snaží rychle vyjádřit myšlenku. Může proto docházet k ovlivňování výslovnosti jednotlivých slov jejich okolím. U spontánní řeči také může mluvčí jednotlivá slova opakovat či některá nedořící do konce, systém pro rozpoznávání pak může daný úsek špatně vyhodnotit.

Na Obrázku 1 je zobrazeno základní schéma rozpoznávání řeči. Mluvčí pronese promluvu, která se nejprve zparametrizuje (viz následující odstavec). Tyto parametry jsou vstupem do systému rozpoznávání, v současné době se pro rozpoznávání používají nejčastěji HMM (viz následující kapitola 3.3 Skryté markovské modely a kapitola 3.4 Princip rozpoznávání řeči pomocí HMM). Ze systému rozpoznávání řeči dostaneme výsledné slovo (v případě rozpoznávání spojité řeči je výsledkem posloupnost slov).



Obrázek 1: Základní schéma rozpoznávání mluvené řeči

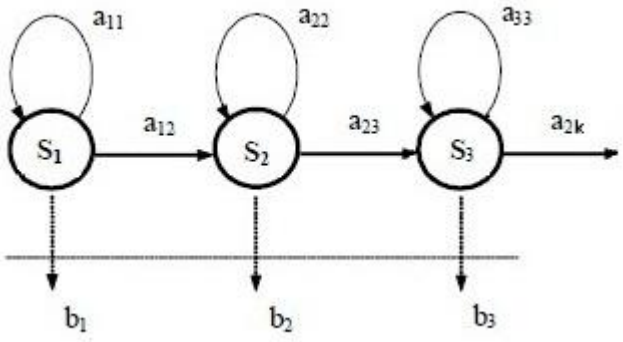
Při rozpoznávání máme k dispozici nahrávku promluvy ve formě posloupnosti jednotlivých vzorků signálu. Pro zpracování je to příliš mnoho redundantních dat, proto se používá parametrizace, která danou sekvenci vzorků převede na menší množství příznaků vhodných k rozpoznávání. Signál se nejprve rozdělí na stejně dlouhé segmenty (v řádu několika milisekund), které se nazývají framy. Na každý frame se aplikuje okénko, nejčastěji Hammingovo, a vypočítá se příznakový vektor. Signál **X** je pak reprezentován časovou

posloupností příznakových vektorů $x_1, x_2, \ldots, x_T$, kde T je počet framů. Velmi často se používá parametrizace pomocí MFCC (Mel-frequency cepstral coefficients - melovské frekvenční kepstrální koeficienty), která vychází z frekvenční oblasti a stejně jako sluch u lidí využívá logaritmickou stupnici. Způsob výpočtu příznakových vektorů není předmětem této práce, dá se nalézt například v [Nouza09].

Při rozpoznávání izolovaných slov představuje blok klasifikace nalezení nejpravděpodobnější položky ze slovníku. Při rozpoznávání spojité řeči se hledá nejpravděpodobnější sekvence slov.

3.3 Skryté markovské modely

Pro každý foném se vytvoří markovský model. Struktura markovského modelu je vidět na Obrázku 2, model je lineární, levo-pravý (přechod je možný pouze z levého do pravého stavu). Každý markovský model reprezentuje krátký, téměř stacionární úsek promluvy (konkrétně foném). Řečové ústrojí člověka se při vyslovování hlásek nepřestaví okamžitě, je proto možné předpokládat, že určitá krátká část promluvy (v řádu milisekund) je vždy stacionární. Modely všech fonémů mají vždy stejný počet stavů, obvykle to bývají tři stavy.



Obrázek 2: Třístavový model reprezentující jeden foném

Akustický signál se při parametrizaci rozdělí do framů. Pro každý frame se vypočítá vektor příznaků. Jeden stav markovského modelu poté reprezentuje sekvenci příznakových vektorů, jejichž hodnoty se mění jen málo. Z hodnot příznakových vektorů se vypočítávají parametry pro jednotlivé stavy modelu. Hodnoty a_{ii}, které udávají pravděpodobnost, že model setrvá v aktuálním stavu, se statisticky zjišťují z počtu framů přiřazených k danému

stavu. Přechodové pravděpodobnosti a_{ii+1} tvoří s pravděpodobnostmi a_{ii} komplementární jev, takže vždy platí rovnost:

$$a_{ii} + a_{ii+1} = 1 \tag{3.1}$$

Hodnoty b_i jsou pravděpodobnostní výstupní hodnoty s vícemodálním Gaussovým rozložením. Jedné složce Gaussova rozložení se říká mixtura. Hodnota $b_i(x)$ vyjadřuje míru pravděpodobnosti, že frame popsaný příznakovým vektorem *x* přísluší stavu *i* (byl jím vygenerován). Vypočte se podle následujícího vztahu:

$$b_i(x) = \sum_{\mathrm{m}=1}^{\mathrm{M}} c_{im} \frac{1}{\sqrt{(2\pi)^P \det \Sigma_{\mathrm{im}}}} \cdot \exp\left[-\frac{1}{2}(x - \bar{x}_{\mathrm{im}})^{\mathrm{T}} \Sigma_{\mathrm{im}}^{-1} (x - \bar{x}_{\mathrm{im}})\right] \tag{3.2}$$

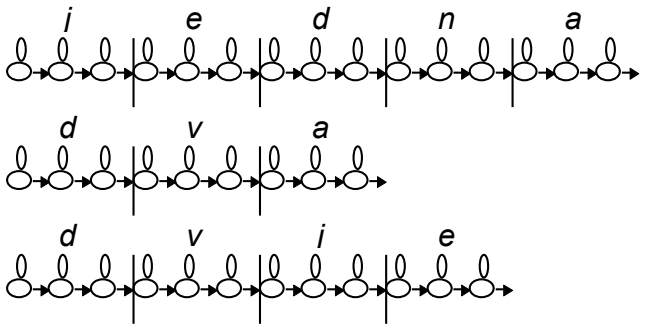
kde c_{im} je váhový koeficient *m*-té mixtury stavu *i* (přičemž platí $\sum_{m=1}^{M} c_{im} = 1$). Symbol Σ značí kovarianční matici a $\bar{x}_{im}$ značí vektor středních hodnot *m*-té mixtury stavu *i*.

Pokud je pro jeden foném vytvořen jeden model, mluvíme o modelu monofonu. Monofon je kontextově nezávislý model hlásky. V případě, že je pro jeden foném vytvořeno více modelů, které jsou závislé na okolních fonémech (konkrétně na fonémech vlevo a vpravo od daného fonému), mluvíme o modelech trifonů. Trifony umožňují lépe modelovat variabilitu fonémů, ale k jejich natrénování je potřeba mnohem větší trénovací sada, která obsahuje všechny modelované trifony.

Počet monofonů se liší podle jazyka. Tento počet je ale relativně malý, což je výhodné, jedná se o 20-50 monofonů. Pro češtinu existuje 40 monofonů. Počet všech trifonů je roven třetí mocnině počtu fonémů. V praxi se však nepoužívají všechny kombinace trifonů, některé se v daném jazyce vyskytují zřídka a jiné se nevyskytují vůbec.

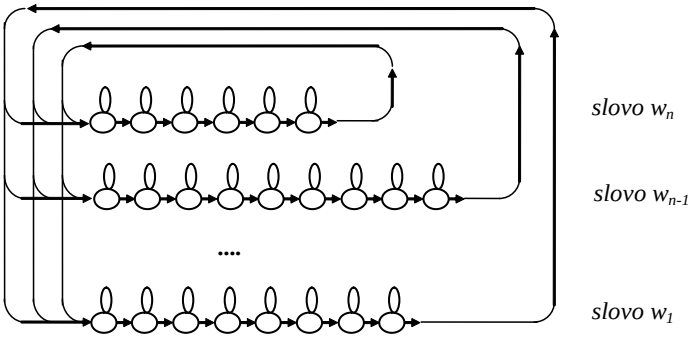
Model slova je vidět na Obrázku 3. Je vytvořen zřetězením modelů jednotlivých fonémů, které jsou předem známy z fonetického přepisu slova.

Význam	Text	Fonetická transkripce
1	jedna	jedna
2	dva	dva
2	dvě	dvje
3	tři	tŘi
4	čtyři	čtiři
5	pět	pjet
6	šest	šest
7	sedm	sedm
7	sedm	sedum



Obrázek 3: Ukázka části slovníku a několika modelů slov, které byly vytvořeny zřetězením modelů fonémů známých z fonetického přepisu (obrázek byl převzat z [Nouza09] se svolením autora)

Model spojité promluvy je vidět na následujícím Obrázku 4. Podle fonetických přepisů slov ze slovníku se vytvoří modely jednotlivých slov. Jazykový model určuje, které slovo může následovat po aktuálním rozpoznaném slově. Jazykové modely jsou převážně dvojího typu: první typ určuje pevnou gramatiku, která udává konkrétní slova, která mohou po aktuálním rozpoznaném slově následovat. Druhý typ jazykového modelu je pravděpodobnostní a využívá se ve většině moderních diktovacích systémů. Po aktuálním slově může následovat jakékoliv další slovo obsažené ve slovníku, každému přechodu je ovšem dána určitá pravděpodobnost. Tyto pravděpodobnosti se často získávají statistickým vyhodnocením velkého množství textu v daném jazyce.



Obrázek 4: Model spojité promluvy (obrázek byl převzat z [Nouza09] se svolením autora)

Akustickým modelem promluvy se myslí soubor modelů jednotlivých fonémů a popřípadě i modelů mimoslovních ruchů.

3.4 Princip rozpoznávání řeči pomocí HMM

Rozpoznávání izolovaných slov spočívá ve vyhodnocení pravděpodobnosti $P(\mathbf{\Phi}|\mathbf{X})$, že neznámé slovo $\mathbf{X}$ patří k některému modelu slova, který je reprezentován množinou parametrů $\mathbf{\Phi} = \{c_{im}, a_{ii}, a_{ii+1}, b_i;\ \forall i \in 1..I, \forall m \in 1..M\}$, kde I je počet stavů slova a M je počet mixtur. Slovo ze slovníku reprezentované parametry modelu $\mathbf{\Phi}$, pro které vyšla pravděpodobnost $P(\mathbf{\Phi}|\mathbf{X})$ nejvyšší, bude prohlášené za výsledné rozpoznané slovo.

Pravděpodobnost $P(\mathbf{\Phi}|\mathbf{X})$ se podle [Nouza09] určí jako maximální pravděpodobnost z hodnot vypočítaných přes všechna přiřazení neznámého slova $\mathbf{X}$ ke stavům modelu daného slova reprezentovaného parametry $\mathbf{\Phi}$. Maximální pravděpodobnost $P(\mathbf{\Phi}|\mathbf{X})$ se dá zjistit pomocí Viterbiho algoritmu. Konkrétní postup Viterbiho algoritmu se dá nalézt také například v [Červa04].

Systémy pro rozpoznávání spojité řeči mají úlohu ještě mnohem složitější. V případě rozpoznávání izolovaných slov jsme věděli, že posloupnost příznakových vektorů $\mathbf{X}$ odpovídá jedné položce ze slovníku. U spojité řeči tomu tak není, mohou se zde vyskytovat různá slova ze slovníku, začínat a končit v libovolném časovém okamžiku. Podrobnější popis rozpoznávání spojité řeči lze nalézt v [Nouza09].

Pro vyhodnocování úspěšnosti rozpoznávání izolovaných slov se používá míra úspěšnosti Corr (Correctness – správnost). Ta v procentech udává, kolik slov bylo rozpoznáno správně. Vypočítá se podle vzorce:

$$Corr = \frac{H}{N} \cdot 100\% \tag{3.3}$$

kde H vyjadřuje počet správně rozpoznaných slov a N je počet všech testovaných slov.

Při vyhodnocování úspěšnosti spojité řeči se míra úspěšnosti Corr využívá také, ale není v tomto případě příliš objektivní. Rozpoznávací systém může chybně rozpoznat slova, která v promluvě vůbec nejsou a protože míra úspěšnosti Corr se vypočítává pouze ze správně rozpoznaných slov a z celkového počtu slov, které jsou obsaženy v promluvě, tak tyto chyby nebere v potaz. Například pokud by mluvčí přečetl větu „Šel jsem do lesa“ a systém by rozpoznal posloupnost slov „Šel jsem sám do lesa“, tak by míra úspěšnosti Corr vyšla rovných 100 %, což ale neznamená, že věta byla rozpoznána správně. Byla proto zavedena další míra úspěšnosti Acc (Accuracy – přesnost), která zohledňuje i slova vložená. Vypočte se podle následujícího vzorce:

$$Acc = \frac{H - I}{N} \cdot 100\% \tag{3.4}$$

kde I udává počet nesprávně vložených slov. Míra Acc je objektivnější s ohledem na celkovou správnost rozpoznání řeči. Míra Corr je vždy kladná, ale míra úspěšnosti Acc může vyjít i v záporných hodnotách.

3.5 Princip trénování modelů HMM

Trénování modelů HMM je nejtěžší úloha markovských modelů. Bylo vytvořeno mnoho různých algoritmů pro trénování modelů, například Baum-Welchův algoritmus (Forward-Backward) nebo také Viterbiho algoritmus. Konkrétně se při trénování určují hodnoty přechodových pravděpodobností, pravděpodobností setrvání ve stavu, výstupní pravděpodobnosti a váhy jednotlivých mixtur. Tyto parametry se určují z testovacích nahrávek a jejich přesných fonetických přepisů.

Trénování parametrů markovských modelů je možné provést pomocí Viterbiho algoritmu, který přiřadí jednotlivé framy trénovaného fonému ke stavům jeho modelu. Jednotlivé framy fonému jsou nejprve rovnoměrně přiřazeny ke stavům daného modelu. Provede se prvotní odhad všech parametrů modelu. Poté se parametry v několika iteracích optimalizují. Tyto optimalizace využívají Viterbiho algoritmu (viz například [Červa07]).

Baum-Welchův algoritmus nepřiřazuje framy pevně k jednotlivým stavům modelu, ale udává pravděpodobnost, se kterou daný frame patří k určitému stavu. Jakýkoliv frame tak může být přiřazen k jakémukoliv stavu modelu.

3.6 Principy adaptace

Úlohy adaptace se dělí podle třech základních parametrů:

1) Fonetický přepis adaptačních dat
   - *Řízená adaptace* – k dispozici je vytvořený fonetický přepis, který považujeme za věrohodný (většinou vytvořený člověkem)

- *Neřízená adaptac*e – fonetický přepis k dispozici není, ale je možné ho vytvořit automaticky vhodným programem (předpokládáme určitou chybovost)

2) Použití adaptačních dat
   - *Postupná adaptace* (inkrementální adaptace) – adaptační data přicházejí do systému postupně a ten se jimi adaptuje tak, jak přicházejí
   - *Dávková adaptace* – všechna adaptační data jsou přístupná najednou a systém se všemi rovnou adaptuje
3) Typ adaptace
   - *Adaptace akustického modelu* – upravují se parametry akustického modelu
   - *Transformace vektoru příznaků* – transformují (normalizují) se přímo příznakové vektory řečových signálů

V rámci této práce byla vždy používána řízená dávková adaptace. Mezi testované metody patřily jak metody adaptace akustického modelu (např. metoda MAP), tak i metody transformace vektoru příznaků (metoda CMLLR).

3.6.1 Metoda MAP

Metoda MAP (Maximum A Posteriori – maximální aposteriorní pravděpodobnost) je založena na odhadu nových parametrů modelů z adaptačních dat metodou maximální aposteriorní pravděpodobnosti. Jako hodnoty parametrů apriorních rozložení jsou většinou využívány přímo odpovídající hodnoty parametrů adaptovaného modelu. Ostatní apriorní parametry mají význam volitelné adaptační váhy. Vztah pro odhad vektoru středních hodnot adaptovaného systému (SA – Speaker Adapted – adaptovaný na mluvčího) lze vyjádřit jako:

$$\widehat{\boldsymbol{\mu}}_{im}^{SA} = \frac{\tau_{im}\boldsymbol{\mu}_{im}^{SI} + \sum_{t=1}^{T} \zeta_t(i,m)\, x_t}{\tau_{im} + \sum_{t=1}^{T} \zeta_t(i,m)} \tag{3.5}$$

kde symbol $\sum_{t=1}^{T} \zeta_t(i,m)$ značí okupační věrohodnost m-té mixtury stavu i daného modelu, která je vypočtena z adaptačních dat $\mathbf{X}$. Symbol $\boldsymbol{\mu}_{im}^{SI}$ značí střední hodnotu původního neadaptovaného (SI – Speaker Independent – nezávislý na mluvčím) modelu a člen τ_{im} má význam adaptační váhy. Pokud by byla hodnota τ_{im} dané komponenty rovna její

celkové okupační věrohodnosti, vyjadřoval by vztah (3.5) odhad středních hodnot této komponenty vypočtený metodou maximální věrohodnosti dohromady z adaptačních i trénovacích promluv. Více o této metodě lze nalézt například v [Červa04].

3.6.2 Metoda MLLR

Metoda MLLR (Maximum Likelihood Linear Regression – maximálně věrohodná lineární regrese) je jedna z nejrozšířenějších metod založených na lineární transformaci. Transformace parametrů je prováděna tak, aby byla maximalizována pravděpodobnost toho, že adaptační data reprezentovaná posloupností příznakových vektorů **X** byla vygenerována daným modelem.

Metoda MLLR se používá především pro transformaci vektorů středních hodnot, která je popsaná následovně:

$$\boldsymbol{\mu}_{im}^{SA} = \mathbf{W}\boldsymbol{\xi}_{im}^{SI} \tag{3.6}$$

kde **W** je hledaná transformační matice a $\boldsymbol{\xi}_{im} = [\omega, \mu_{im1}, \mu_{im2}, \dots, \mu_{imp}]^{\mathrm{T}}$ je rozšířený vektor středních hodnot *m*-té mixtury *i*-tého stavu modelu s posunutím ω.

Mezi největší výhody metody MLLR patří skutečnost, že pro adaptaci stačí menší množství adaptačních dat. Je to dáno tím, že jedna transformační matice může být shodná pro více Gaussových komponent systému, které jsou obsaženy v jedné regresní třídě. Způsobů, jak zjistit, které komponenty budou v jedné regresní třídě je více. V prostředí HTK se využívá algoritmus klastrování, který vytvoří takzvaný regresní strom. Každý uzel tohoto binárního stromu reprezentuje skupinu akusticky podobných komponent, viz [Young09].

3.6.3 Metoda CMLLR

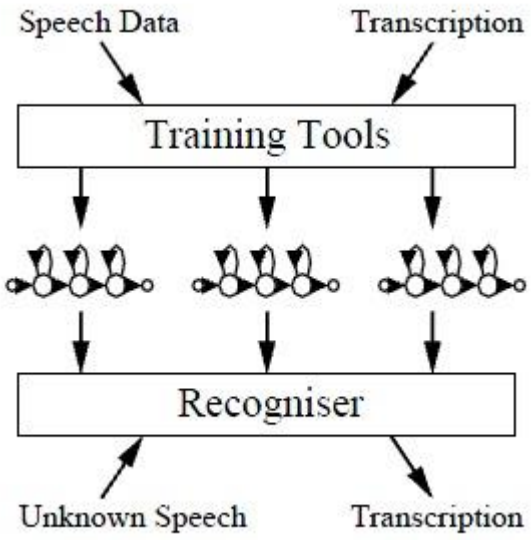
Metoda CMLLR (Constrained maximum likelihood linear regression – omezená maximálně věrohodná lineární regrese) je založena na metodě MLLR. Snaží se najít transformaci, která snižuje rozdíly mezi trénovací sadou a adaptačními daty. CMLLR je přesněji funkce, která odhaduje sadu lineárních transformací pro adaptaci příznakových vektorů. Transformace se snaží změnit příznakové vektory trénovací sady tak, aby byla co největší pravděpodobnost, že každý markovský stav systému vygeneroval adaptační data.

Společně s metodou CMLLR se využívá adaptační technika SAT (Speaker Adaptive Training – trénování pro účely adaptace na mluvčího). Tato metoda se používá pro vytváření modelů nezávislých na mluvčím, navržených přímo pro adaptaci. Cílem této metody je potlačit rozdílnost mezi jednotlivými řečníky a vytvořit tak přesnější akustické modely s menšími hodnotami rozptylů.

4. Prostředí HTK

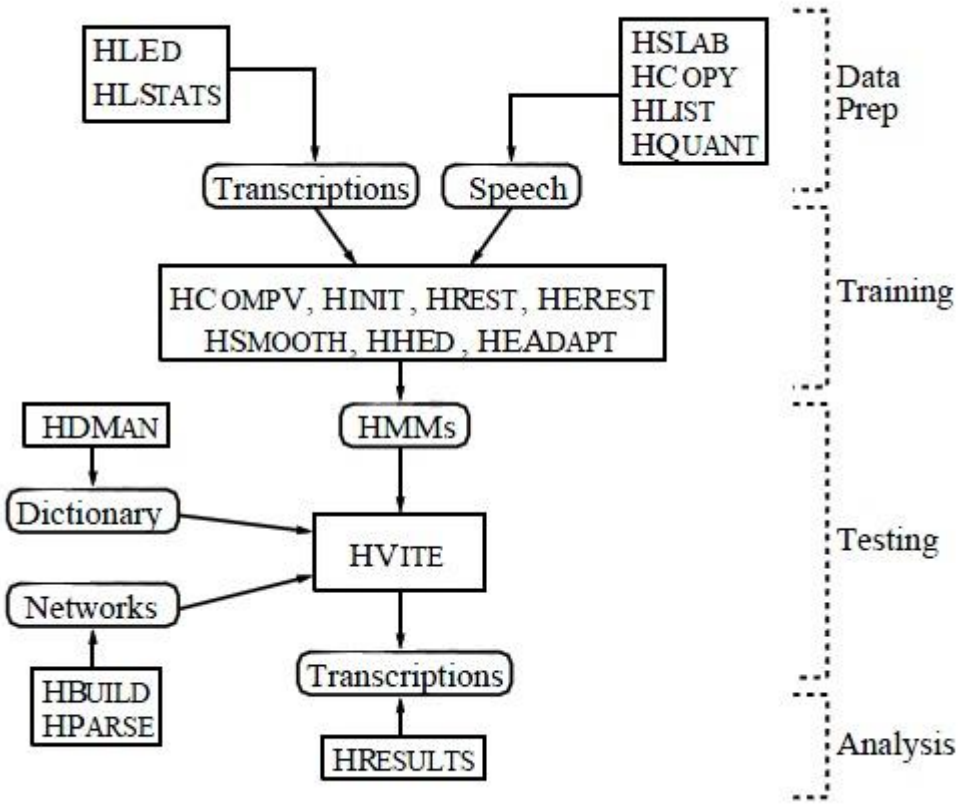
HTK je sada nástrojů pro vytváření skrytých markovských modelů, které vyvinula Speech Group z University v Cambridge. HTK bylo primárně vyvinuto pro vytváření markovských modelů, které jsou určeny pro práci s řečovými signály, zejména pro rozpoznávání řeči. Markovské modely je ale možné použít pro modelování jakékoliv časové řady a jádro HTK je obecně účelové a použitelné i pro jiné aplikace, jako je například rozpoznávání znaků nebo sekvenční zpracování DNA.

Obrázek 5 zobrazuje hlavní strukturu HTK, která se skládá ze dvou komponent. První komponenta obsahuje různé nástroje pro odhad parametrů sady HMM za použití trénovacích promluv a jejich přesných fonetických přepisů. Druhá komponenta zahrnuje nástroje pro rozpoznání a přepis neznámé promluvy.



Obrázek 5: Schéma základní struktury nástrojů poskytovaných HTK (Obrázek byl převzat z [Young09])

Nástroje obsažené v HTK je možné rozdělit do 4 hlavních skupin: nástroje pro přípravu dat, nástroje pro trénování, nástroje pro samotné rozpoznávání a nástroje pro analýzu. Následující Obrázek 6 zobrazuje jednotlivé nástroje a jejich zařazení do zmíněných skupin.



Obrázek 6: Struktura jednotlivých nástrojů HTK a jejich zařazení do 4 hlavních skupin (Obrázek byl převzat z [Young09])

Před samotným použitím se zvukové nahrávky nejprve zparametrizují, například do formátu MFCC. Přepisy daných nahrávek musejí být ve správném formátu. Některé z nejdůležitějších nástrojů pro tuto práci v HTK jsou následující:

- HLEd – tento nástroj slouží k práci se soubory s fonetickým přepisem. Používá se například pro transformaci původního souboru do formátu používaného ostatními nástroji HTK.
- HLStats – používá se pro shromažďování a zobrazování statistik o souborech s přepsanými zvukovými nahrávkami
- HSLab – používá se pro ruční vytváření a úpravu anotačních souborů pro jednotlivé nahrávky
- HCopy – tento nástroj se využívá při parametrizaci vstupních zvukových záznamů, parametrizace transformuje vstupní nahrávky na příznakové vektory MFCC

- HList – využívá se pro kontrolu obsahu zvukových souborů a pro kontrolu výsledků konverze před zpracováním velkého množství dat

Další sada nástrojů vytváří topologii potřebnou pro každý HMM, prototyp pro inicializaci a následné odhady modelů. Nejdůležitější nástroje této části jsou:

- HCompV – nástroj pro výpočet středních hodnot a rozptylů pro inicializaci modelů
- HInit – používá se pro výpočet počátečních parametrů pomocí fonetických přepisů trénovacích nahrávek. Tento nástroj využívá Viterbiho algoritmus, který hledá nejpravděpodobnější sekvenci stavů odpovídající každému trénovacímu vzorku.
- HRest – nástroj pro přetrénování parametrů modelů za pomoci Baum-Welchova algoritmu. Používá mluvenou promluvu jako zdroj trénovacích dat a vytváří z nich kompletní sadu HMM. Pro každou promluvu je potřeba mít přesné fonetické přepisy, přesněji sekvenci fonémů v dané promluvě.
- HHed – editor HMM, umožňuje editaci HMM struktur, jako je například klonování modelů nebo navyšování počtu mixtur
- HParse – ze zadané gramatiky vytvoří soubor, který obsahuje síť udávající přípustné sekvence slov. Tento soubor je používán v dalším bloku nástrojů HTK

Další skupina nástrojů se používá pro samotné rozpoznávání řeči. Umožňuje práci v režimu online či offline:

- HVite – tento nástroj umožňuje rozpoznávání řeči pomocí jazykových a akustických modelů za využití Viterbiho algoritmu. Vstupem je vytvořená síť určující povolenou sekvenci slov vytvořená nástrojem HParse, slovník definující výslovnost každého slova (je možné uvést pro jedno slovo i více výslovností) a sada HMM. Rozpoznávání může být spuštěno nad sadou předem vytvořených nahrávek či nad živým nahráváním zvuku. Jedním z velmi důležitých parametrů tohoto nástroje je Penále. To udává fixní hodnotu, která se ke každému slovu přičte. V praktickém použití to znamená, že jedno delší slovo má při rozpoznávání větší pravděpodobnost než několik krátkých slov.

Poslední sada nástrojů zahrnuje programy pro analýzu, které vyhodnocují úspěšnost rozpoznávání. Obvykle se úspěšnost zjišťuje přes porovnání přepsaných testovacích dat a výsledků ze samotného rozpoznávání. Nástroj HResult je vhodný pro tyto účely:

- HResults – pomocí dynamického programování provádí porovnání správných přepisů a rozpoznaných slov a vrací míru úspěšnosti Corr a Acc

Popis všech nástrojů a způsob jejich použití je podrobně popsán v [Young09].

5. Databáze nahrávek

Aby se dalo objektivně zjistit, zda je úspěšnost rozpoznávání závislá na konkrétním mikrofonu, bylo potřeba vytvořit vhodnou databázi nahrávek. Použití databáze s jednotlivými nahrávkami pořízenými od různých lidí a z různých mikrofonů by nebylo pro tyto účely vhodné, protože každé dvě nahrávky se liší. I když mluvčí pronese do dvou nahrávek stejnou promluvu, tak se nahrávky mohou lišit. Tato změna je dána rozdílným hlukem v pozadí, mluvčí může pronést promluvu jinou rychlostí, či se může nadechnout v jiném místě a podobně. Proto byla, pro účely této diplomové práce, vytvořena databáze „paralelních" nahrávek nazvaná MultiMic Database. Databáze obsahuje nahrávky od různých mluvčích, přičemž v jednu chvíli snímaly mluvčího dva mikrofony zároveň. Jeden mikrofon byl vždy referenční. Dosáhlo se tak tím toho, že takové dvě nahrávky se již dají mezi sebou porovnávat.

Pro vytvoření databáze bylo k dispozici 6 různých sluchátek s mikrofonem (viz Tabulka 1). Jako referenční mikrofon byl vybrán model sluchátek s mikrofonem HS-741 od výrobce i-Tec, které měl mluvčí umístěn kolem krku. Druhá sluchátka s mikrofonem, který zaznamenával promluvu, byla umístěna na hlavě mluvčího (viz Obrázek 7). Vytvořená databáze celkem obsahuje nahrávky od 6 různých osob – 3 žen a 3 mužů. Každý z mluvčích nahrál 5 různých nahrávek do dvou mikrofonů zároveň. Databáze se tedy skládá z celkem 60 různých nahrávek, které dohromady trvají více než 1 hodinu.

VÝROBCE	TYP
i-Tec	**HS-741**
Genius	HS-02N
KOSS	CS/100
Sennheiser	PC 111
Sennheiser	PC 131
Sennheiser	PC 161

Tabulka 1: Použité typy sluchátek s mikrofonem - zvýrazněný mikrofon byl použit jako referenční

Obrázek 7: Mluvčí nahrávající do dvou mikrofonů zároveň

Pro testování vlivu mikrofonů na rozpoznávání (viz Kapitola 6.1.1 Vliv mikrofonu na rozpoznávání) se nahrávky použily v podobě, ve které byly nahrány. Každá nahrávka z vytvořené databáze se skládá ze dvou odlišných částí, pro další testování (adaptačních metod) se nahrávka ručně rozdělila na dvě části. Mezi těmito dvěma částmi nahrávky mluvčí dělali pauzu (2-3 sekundy ticha) dostatečně dlouhou na to, aby po rozdělení nahrávky nebylo rozpoznávání ovlivněno. První část, která se později při testování adaptačních metod používá samostatně jako adaptační část, obsahuje uměle vytvořenou promluvu, která je pro všechny nahrávky stejná (viz Příloha A). Uměle byla vytvořena z toho důvodu, aby obsahovala všechny fonémy a nebyly tak příliš omezeny některé metody adaptace. Tato část obsahuje celkem 63 slov a v průměru trvá 31 sekund. Druhá část nahrávek obsahuje promluvu vytvořenou z dostupných článků, které zajistily různorodost (ukázka jedné promluvy je v Příloze A). Tato promluva byla pro všechny nahrávky jiná. Druhá část nahrávky byla v pozdějších testech samostatně používána jako testovací část, na které se zkoumalo, jaké jsou výsledky adaptace. Počet slov ve druhé části se pohybuje mezi 50 a 100 slovy (v průměru se jedná o 70 slov na jednu testovací nahrávku). Průměrná délka testovací části je 33 sekund. Vzhledem k tomu, že počet slov testovací části byl ve většině případů vyšší než počet slov adaptační části, měla by testovací část trvat delší dobu. To odpovídá zjištěným hodnotám, ale

rozdíl průměrných délek je o něco nižší než původní předpoklad. Je to dáno jednak tím, že slova jsou různě dlouhá a také tím, že adaptační část se v rámci jedné nahrávky nahrávala jako první a mluvčí ze začátku promluvy dával větší pozor na správné vyslovování a ke konci se jeho promluva o něco zrychlila.

Pro adaptační nahrávky bylo potřeba vytvořit soubory s fonetickým přepisem. Tyto přepisy byly vytvořeny ručně při poslechu každé nahrávky, aby byla zajištěna co největší přesnost přepisů. Pro testovací nahrávky se vytvořily soubory obsahující na jednom řádku jedno slovo z dané promluvy. Tyto soubory byly použity pro zjištění míry úspěšnosti.

Druhá testovací databáze byla vytvořena za účelem otestování vlivu polohy mikrofonu před ústy. Při nahrávání MultiMic Database se nahrávalo do dvou mikrofonů zároveň. Nahrávalo se, jak již bylo výše zmíněno tak, že jedna sluchátka měl mluvčí umístěna na hlavě a mikrofon byl standardně vyveden před ústa, druhá sluchátka měl mluvčí kolem krku a mikrofon byl nastaven do pozice před ústy. Tato databáze obsahuje nahrávky z jednoho (referenčního) mikrofonu v těchto dvou pozicích před ústy, aby bylo možné otestovat, zda je rozdíl v úspěšnostech mezi mikrofony ovlivněn polohou mikrofonu před ústy. Celkem bylo nahráno 199 krátkých nahrávek (5 až 20 slov na jednu nahrávku). Polovina nahrávek byla vytvořena se sluchátky na hlavě, druhá se sluchátky kolem krku. Nahrávky byly rozděleny do osmi bloků, každý blok se nahrával samostatně, s minimálním časovým odstupem 2 hodin. Dosáhlo se tak určité variability nahrávek a zaručilo se tak, že mikrofon byl pokaždé v trochu odlišné pozici před ústy.

Pro testování základních metod rozpoznávání řeči a vlivu jejich parametrů, byla vytvořena databáze nahrávek, která obsahuje izolovaná slova. Tato databáze byla pořízena jedním mluvčím, obsahuje celkem 110 různých, izolovaně pronesených slov. Tato slova byla záměrně vybrána velmi podobná (např.: Míla, bílá, byla, milá, lila, vila, apod.), aby byly dobře patrné rozdíly výsledné úspěšnosti pro odlišná nastavení vstupních parametrů.

Aby bylo možné otestovat základní metody pro rozpoznávání a intuitivní adaptační metodu, bylo potřeba vytvořit sadu i pro trénování modelů. Databáze trénovacích nahrávek se skládala jednak z nahrávek pořízených studenty v rámci magisterského předmětu Počítačové zpracování řeči (konkrétně se jednalo o 11 různých mluvčí, 2 ženy a 9 mužů) a jednak z nahrávek pořízených od 9 dalších, různých žen. Každý mluvčí nahrál přibližně 100 nahrávek (nahrávka vždy obsahovala jednu namluvenou větu), které byly rozděleny do dvou částí. Prvních 80 nahrávek obsahovalo vždy jednu větu z dostupných článků. Zbylých 20 nahrávek obsahovalo takové namluvené věty, které doplňovaly celkový počet výskytů méně

častých fonémů. Dosáhlo se tak tím toho, že v každé sadě 100 nahrávek byly všechny fonémy alespoň pětkrát. Trénovací databáze obsahuje celkem 1982 nahrávek od 20 různých osob.

6. Experimentální práce

Cílem experimentů bylo zjistit, zda je počítačové rozpoznávání řeči závislé na použitém mikrofonu. A pokud tomu tak je, jak je možné tuto závislost kompenzovat. První testy byly provedeny na systému rozpoznávání řeči vytvořeném v Laboratoři počítačového zpracování řeči. Testovaly se nahrávky z vytvořené databáze a zjišťovalo se, jestli jsou úspěšnosti rozpoznávání nahrávek z různých mikrofonů odlišné. Ještě před samotným testováním jednotlivých metod pro adaptaci se ověřily základní metody pro rozpoznávání a byly otestovány různé vstupní parametry. Poté se testovaly různé metody adaptace. Nejprve se vyzkoušela intuitivní metoda, poté se testovaly pokročilejší metody adaptace implementované do programu HTK a nakonec byla otestována adaptace pokročilejších metod pomocí programu na rozpoznávání řeči vyvíjeném v Laboratoři počítačového zpracování řeči.

6.1 Základní experimenty

Cílem těchto experimentů bylo zjistit, zda má mikrofon, který byl použitý při nahrávání, vliv na počítačové rozpoznání řeči a jak je tento vliv výrazný. Součástí prvotních experimentů také bylo ověření metod pro rozpoznávání řeči a seznámení se s jejich použitím v programu HTK. Byly testovány různé vstupní parametry, které ovlivňují úspěšnost rozpoznávání, jako například kvalita přepisů nahrávek, ze kterých se trénovaly modely jednotlivých fonémů, nebo nastavení penalizace při rozpoznávání neznámých nahrávek.

6.1.1 Vliv mikrofonu na rozpoznávání

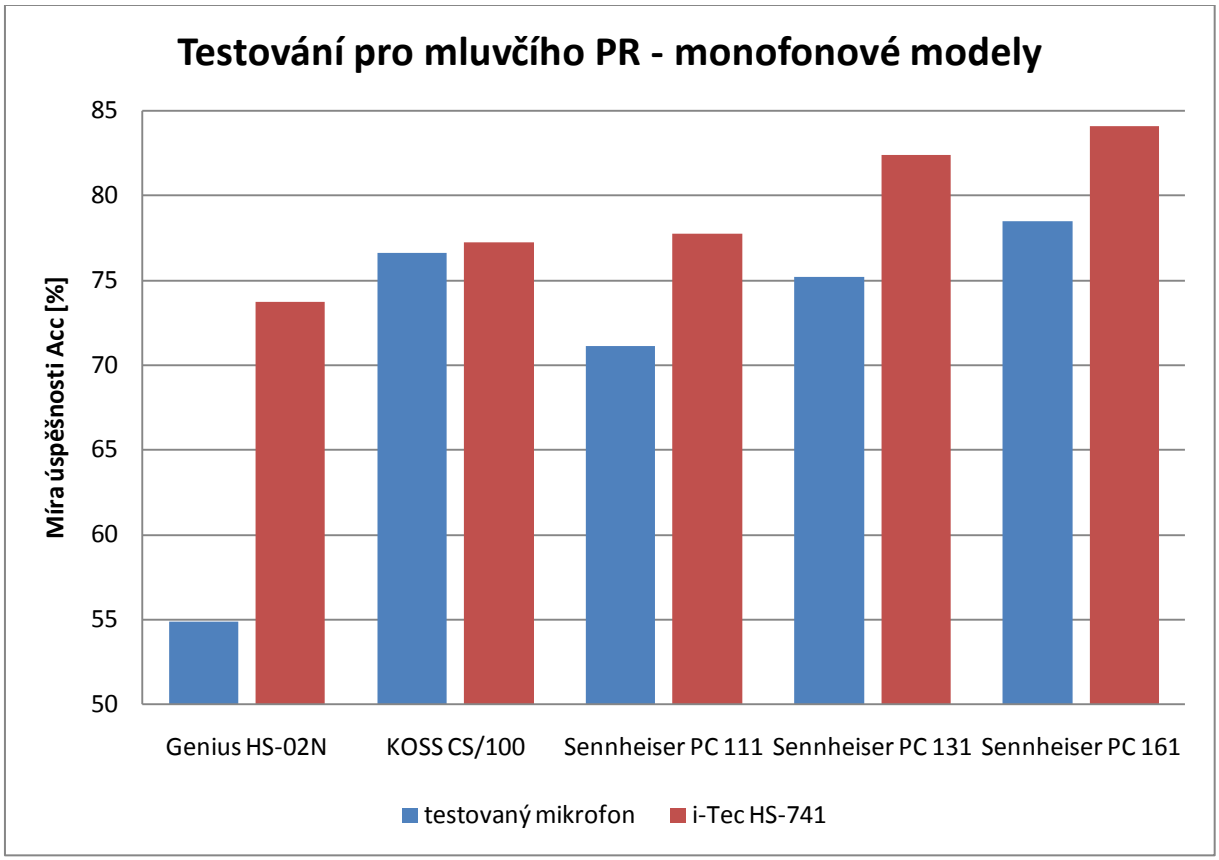
Nejprve bylo potřeba zjistit, zda má použitý mikrofon vliv na úspěšnost rozpoznávání řeči. Testování bylo provedeno na vytvořené databázi MultiMic Database. Všechny nahrávky byly použity v původní podobě (nebyly rozděleny na adaptační a testovací část). Toto testování proběhlo v Laboratoři počítačového zpracování řeči na programu rozpoznávání řeči, který se používá v praktických aplikacích. Testování proběhlo dvakrát, jednou se sadou monofonů, podruhé se sadou trifonů. Druhé testování (se sadou trifonů) proběhlo o necelý rok

později, takže použitý model byl natrénovaný na větší sadě nahrávek. Výsledný rozdíl v úspěšnostech mezi jednotlivými modely je tedy dán nejenom přechodem z monofonů na trifony, ale také tím, že trénovací sada byla ve druhém případě obsáhlejší.

Grafy v této podkapitole (konkrétně se jedná o grafy na Obrázku 8, Obrázku 9, Obrázku 10 a Obrázku 11) jsou vytvořeny se stejným měřítkem (u grafů zobrazujících výsledky rozpoznávání s modely trifonů je osa znázorňující úspěšnost pouze posunuta o 10 % výše, aby byly dobře vidět výsledné hodnoty dosahující vyšších hodnot), aby bylo možné je navzájem porovnávat. Testování proběhlo na stejných systémech a odlišnosti jsou pouze v modelech (modely monofonů a trifonů) a v tom, zda jsou výsledky uvedeny pro jednu konkrétní osobu či jsou zprůměrovány přes všechny osoby. Konkrétní osoba byla vybrána stejná pro testování s monofony i trifony, aby bylo možné porovnat rozdíl mezi modely na stejných nahrávkách.

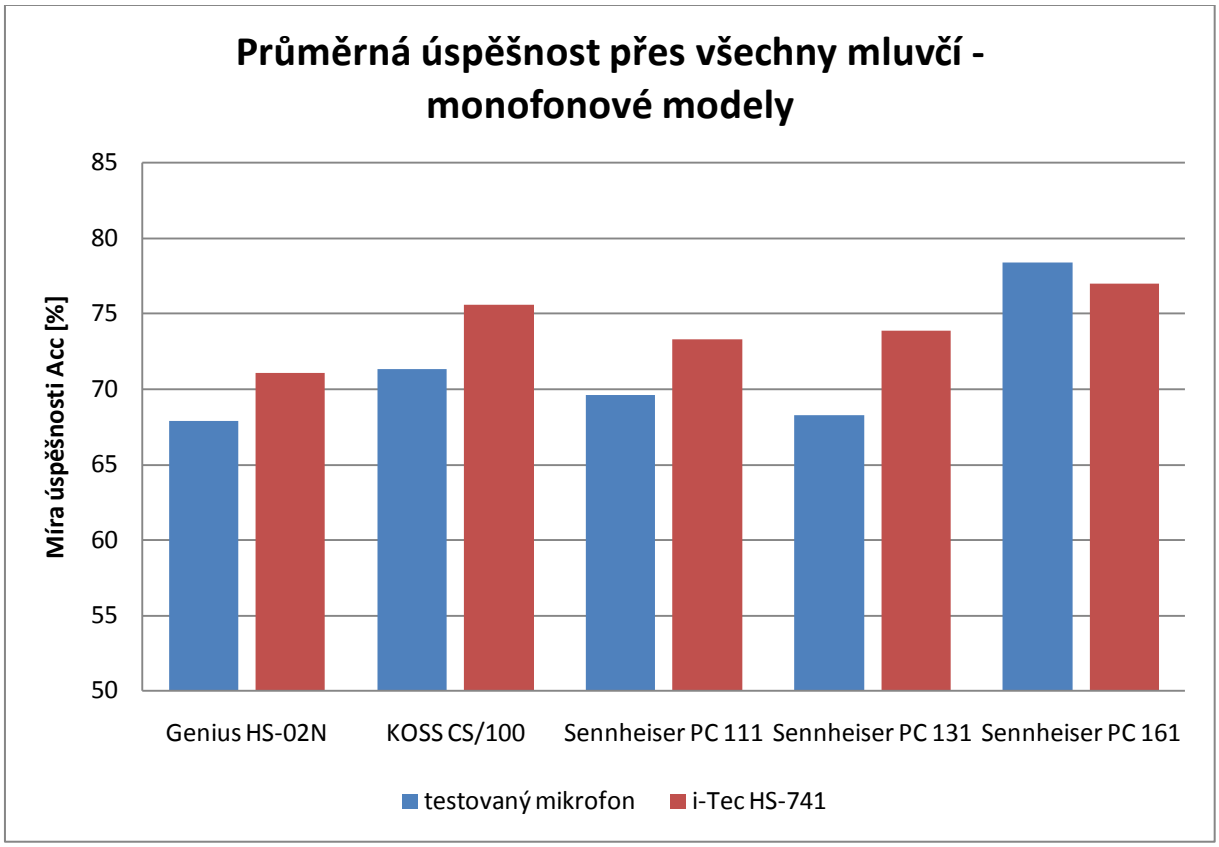
Výsledky v uvedených grafech jsou v míře úspěšnosti Acc. Tyto výsledky nedosahují očekávaných hodnot úspěšnosti. Je to dáno především volbou první (adaptační) části nahrávek z použité databáze. Tyto nahrávky jsou primárně určeny pro trénování a obsahují všechny fonémy vyskytující se v českém jazyce. Ovšem díky netypické struktuře vět jim pravděpodobnostní jazykový model používaný v systému pro rozpoznávání přiřadil menší pravděpodobnosti. Věty použité pro promluvy adaptační části jsou zobrazeny v Příloze A.

V grafu na Obrázku 8 jsou zobrazeny výsledky z prvního testování pro jednoho konkrétního mluvčího (tabulka výsledků pro všechny mluvčí je v Příloze B) označeného iniciály PR, za použití modelů monofonů. Modrý a červený sloupec, které jsou těsně vedle sebe, znázorňují výsledky úspěšnosti pro jednu nahrávku. Modrý sloupec zobrazuje výslednou úspěšnost rozpoznávání pro zvukovou stopu nahrávanou mikrofonem, který je uvedený pod sloupcem. Červený sloupec zobrazuje výslednou úspěšnost pro zvukovou stopu nahrávanou referenčním mikrofonem (i-Tec HS-741). Z grafu je vidět, že výsledky se v určitých případech velmi liší. Pro tohoto konkrétního mluvčího byl nejvýraznější rozdíl při nahrávání do mikrofonu značky Genius a referenčního mikrofonu. Tento rozdíl přesahuje 18 %. Naopak rozdíl mezi nahrávkami pořízenými referenčním mikrofonem a mikrofonem značky KOSS CS/100 je velmi malý, dosahuje hodnoty přibližně 0,7 %. Pro zbylé tři mikrofony dosahuje rozdíl v průměru 6,5 %.



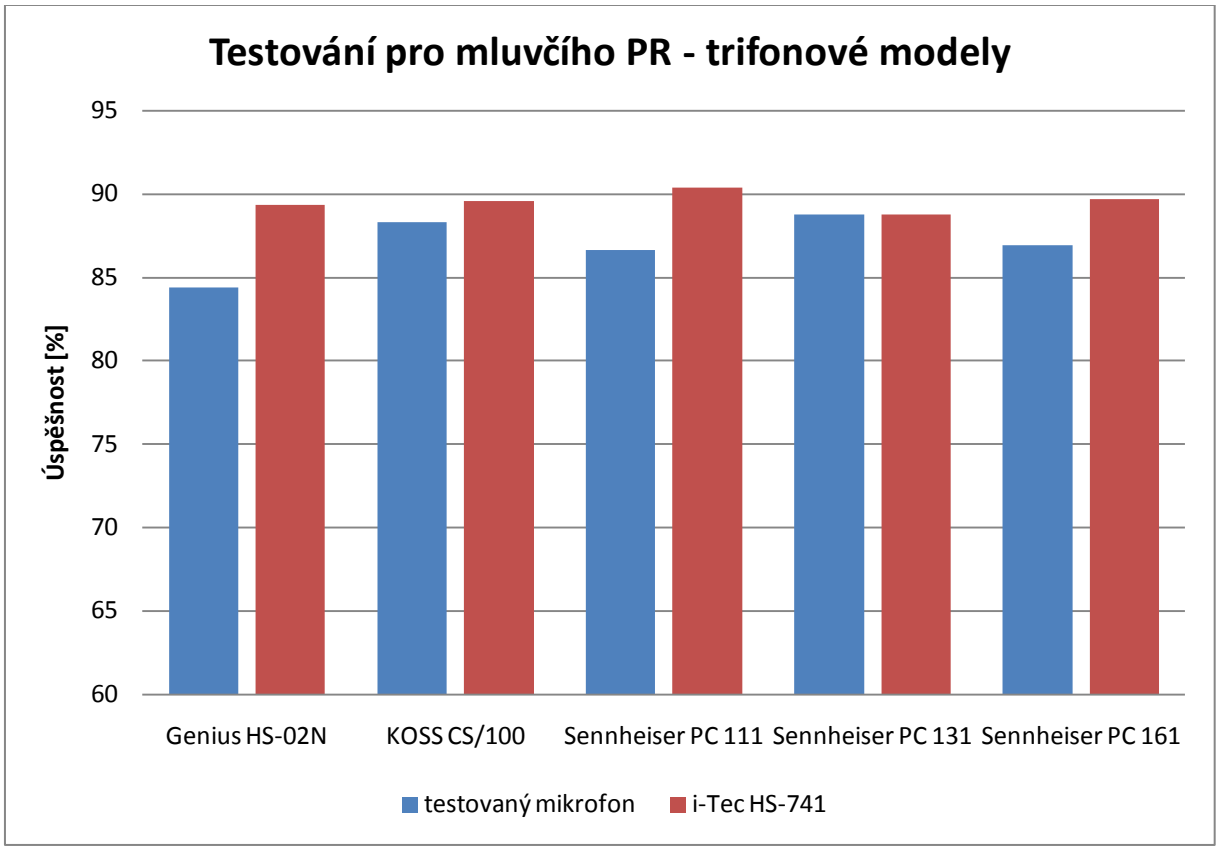
Obrázek 8: Výsledky testování pro jednoho konkrétního mluvčího s monofony

V následujícím grafu znázorněném na Obrázku 9 jsou vidět výsledky testování pro jednotlivé mikrofony zprůměrované přes všechny mluvčí. Z grafu je vidět, že v průměru nedosahují rozdíly mezi jednotlivými mikrofony tak velkých hodnot, jako tomu bylo u výsledků pro konkrétní osobu, znázorněných v grafu na Obrázku 8. Maximální rozdíl je v průměru 5 %. Z grafu je také vidět, že mikrofon, který byl zvolen jako referenční, v průměru dosahuje nejlepších výsledků.



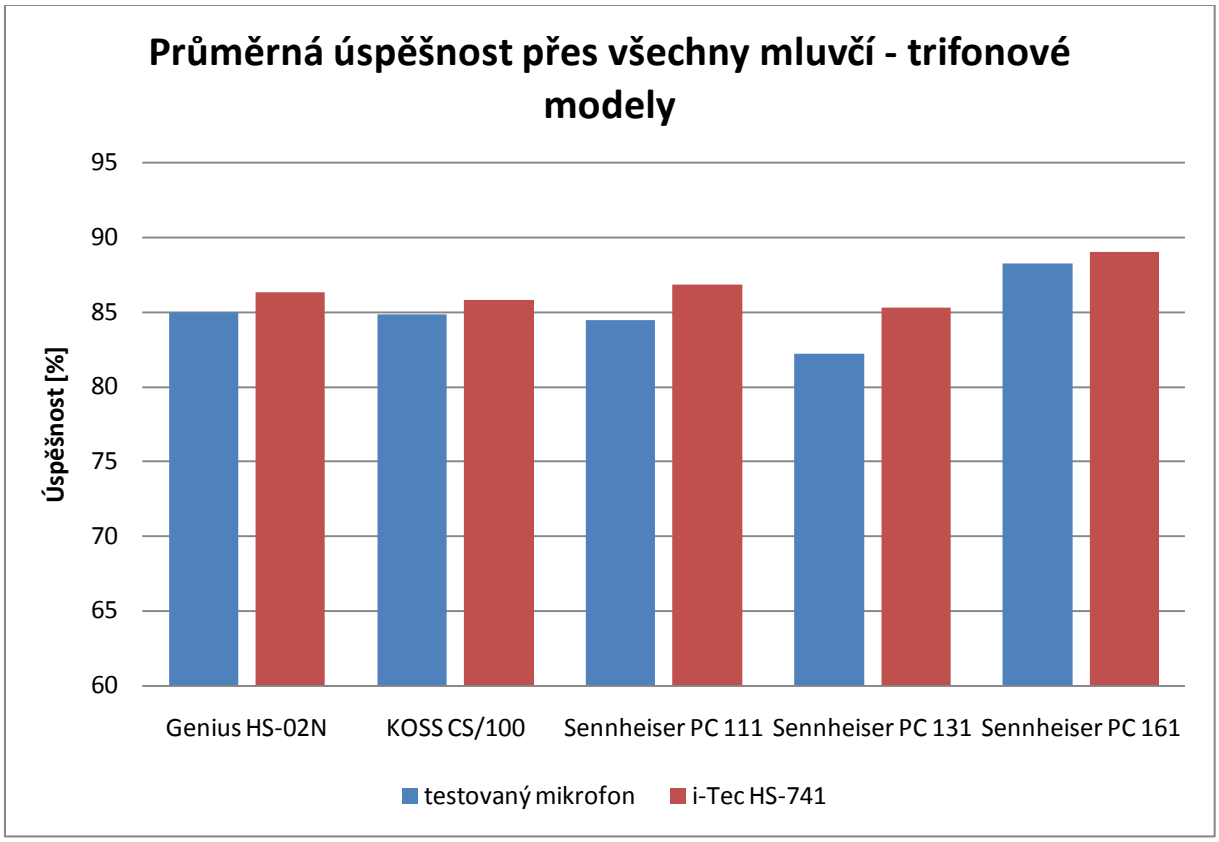
Obrázek 9: Výsledky testování zprůměrované přes všechny mluvčí s monofony

V grafu na Obrázku 10 jsou výsledky testování pro jednoho konkrétního mluvčího s modely trifonů (tabulka výsledků pro všechny mluvčí je v Příloze B). Na první pohled je zřejmé, že výsledné úspěšnosti jsou pro všechny nahrávky lepší. Jak již bylo výše zmíněno, není to pouze přechodem od monofonů k trifonům, ale také zvětšením trénovací sady. Rozdíly mezi mikrofony pro jednotlivé nahrávky jsou menší, ale přesto stále výrazné. Například rozdíl mezi referenční nahrávkou a nahrávkou pořízenou mikrofonem Genius HS-02N byl při použití monofonů vyšší než 18 %. Při použití nového modelu trifonů je tento rozdíl necelých 5 %. Rozdíly u ostatních mikrofonů jsou menší než 4 % a rozdíl mezi referenčním mikrofonem a mikrofonem Sennheiser PC 131 je dokonce nulový.



Obrázek 10: Výsledky testování pro konkrétního mluvčího s trifony

V dalším grafu na Obrázku 11 jsou výsledky testování s modely trifonů, zprůměrované přes všechny mluvčí. Z grafu je vidět, že průměrná úspěšnost je pro většinu mikrofonů o něco nižší než pro uvedeného konkrétního mluvčího. Ale například rozdíl úspěšností pro mikrofon značky Genius HS-02N a referenční mikrofon je o něco menší. Naopak výsledná úspěšnost mikrofonu Sennheiser PC 131 je v průměru horší než ostatní mikrofony, rozdíl činí 3,1 %. V tomto testování se potvrdilo, že referenční mikrofon dosahuje lepších výsledků než ostatní testované mikrofony.



Obrázek 11: Výsledky testování zprůměrované přes všechny mluvčí s trifony

Jednou z hypotéz, proč jsou výsledné úspěšnosti pro nahrávky ze dvou mikrofonů pro jednu promluvu tak rozdílné, byla možnost vlivu polohy mikrofonu před ústy. K otestování této hypotézy byla vytvořena databáze nahrávek pořízených ve dvou různých polohách, které byly použity při nahrávání první databáze MultiMic Database. Testování opět proběhlo v Laboratoři počítačového zpracování řeči za použití nejlepšího natrénovaného modelu (model trifonů).

Výsledky z tohoto testování jsou uvedeny v Příloze C. Průměrný výsledek (míra úspěšnosti Accuracy) je pro první polohu, kdy byla sluchátka kolem krku, 90,6 %. Pro druhou polohu, kdy se sluchátka nacházela na hlavě mluvčího, byla průměrná výsledná úspěšnost rovna 91,1 %. Rozdíl v úspěšnostech je menší než jedno procento. Tento rozdíl je minimální, můžeme z něj tedy usoudit, že různé polohy mikrofonu před ústy při nahrávání databáze nemají na úspěšnost rozpoznávání veliký vliv.

6.1.2 Úvodní experimenty pro ověření dílčích metod

První testování bylo zaměřeno na ověření základních metod a jejich vstupních parametrů. Nejprve byla otestována důležitost kvality fonetických přepisů nahrávek pro trénování modelů. Testoval se také počet iterací při trénování modelů a počet mixtur výstupních funkcí modelů.

Nahrávky určené pro trénování modelu bylo potřeba foneticky přepsat. Pro testování kvality přepisů byly vytvořeny dvě různé sady. První sada, označená jako Approx Train (approximation – přibližný, train - trénovat) byla z části vytvořena studenty v rámci předmětu Počítačové zpracování řeči. Druhá část přepisů byla vytvořena za pomoci programu, který obsahoval základní pravidla pro fonetickou transkripci. Druhá sada, Prec Train (precise - přesný) byla vytvořena v Laboratoři počítačového zpracování řeči za pomoci speciálního programu a ručně překontrolována a upravena.

Pro testování byl vytvořen systém pro rozpoznávání izolovaných slov. Modely monofonů byly vytvořeny z databáze nahrávek Approx Train a z databáze Prec Train. Pro testování se použily nahrávky z databáze izolovaných slov, která obsahovala 110 různých, ale velmi podobných slov. Slovník obsahoval slova z testovacích nahrávek. Gramatika tohoto systému pro rozpoznávání izolovaných slov umožňovala pouze slova ze slovníku a mezi nimi mohlo být ticho. Tato gramatika byla označena jako G1. Výsledky z prvního testování jsou uvedeny v Tabulce 2. Model byl natrénován s různým počtem mixtur výstupní funkce a byl použit různý počet iterací. U vyššího počtu iterací je možné, že se model přetrénuje a výsledky rozpoznávání budou horší. To se v tomto případě ale nepotvrdilo. S vyšším počtem mixtur roste úspěšnost rozpoznávání, ale také čas potřebný k natrénování modelu. Byly proto zvoleny čtyři hodnoty mixtur výstupních funkcí (1, 2, 4 a 8 mixtur).

počet iterací	počet mixtur	Prec Train		Approx Train	
		Corr [%]	Acc [%]	Corr [%]	Acc [%]
5	1	39,6	-49,6	32,4	-64,9
	2	47,8	-24,3	44,1	-33,3
	4	53,2	-0,9	50,5	-21,6
	8	64,9	30,6	56,7	-1,8
10	1	41,4	-35,1	36,9	-53,2
	2	50,5	-9,9	46,9	-34,8
	4	60,4	13,5	50,5	-15,3
	8	**70,3**	**45,1**	**58,6**	**0,9**

Tabulka 2: Výsledky testování izolovaných slov s 1. typem gramatiky

V uvedené tabulce je vidět, že výsledné rozpoznávání se pro různý počet mixtur velmi liší. Rozdíly mezi výsledky jsou větší u míry úspěšnosti Acc. Pro míru úspěšnosti Corr je průměrný rozdíl mezi trénovacími sadami 6,4 % a pro míru úspěšnosti Acc to je 24,2 %. Maximální rozdíl mezi trénovacími sadami je až 44 % (pro míru úspěšnosti Acc, 10 iterací, 8 mixtur).

U druhého testování byl nejprve upraven slovník, přidáním několika neřečových ruchů. Celkem jich bylo použito šest. První neřečový ruch reprezentuje ráz před samohláskou. Tento ráz se nejčastěji objevuje na začátku věty či po pauze, když je první vyřčený foném samohláska. Vzniká tím, že hlasivky se musí nejprve „nastartovat“. Druhý ruch reprezentuje krátký hluk. Jedná se o kliknutí myši, krátké cvaknutí či o mlasknutí, vznikající při odlepování rtů od sebe na začátku promluvy. Další hluk se objevuje zřídka a je podobný předchozímu krátkému cvaknutí, ale je o poznání tišší. Čtvrtý neřečový hluk reprezentuje nadechnutí či vydechnutí. Předposlední neřečový ruch reprezentuje déle trvající hlasitý zvuk, jako je například projíždějící auto pod oknem, hudba, zvonění telefonu a podobně. Jako poslední hluk byl použit váhavý zvuk (hmm, ehm). V trénovací sadě Prec Train byly tyto hluky uvažovány již při přepisu a byly tedy rovnou obsaženy ve fonetické transkripci. U druhé sady trénovacích nahrávek, Approx Train, byly tyto hluky při přepisu studenty přidávány zřídka. Při strojovém přepisu nebyly přidávány vůbec, ale později byly přidány při ruční kontrole.

Gramatika pro druhé testování (označená jako G2) byla upravena tak, že se před a za nalezené slovo povolila možnost některého z uvedených ruchů. V Tabulce 3 jsou vidět výsledky tohoto testování. Opět se testoval počet iterací a počet mixtur výstupních funkcí. Výsledné úspěšnosti (v míře Corr) jsou oproti předchozí gramatice G1 pro sadu Prec Train vyšší v průměru o 2,1 %, kdežto sada Approx Train je vyšší pouze o 1,7 %. Pro míru

úspěšnosti Acc jsou tyto rozdíly větší. Sada Prec Train se zlepšila v průměru o 7,2 % a sada Approx Train se zlepšila o 4,6 %. Tyto rozdíly mezi sadami jsou dány především již zmíněným postupem při transkripci.

počet iterací	počet mixtur	Prec Train		Approx Train	
		Corr [%]	Acc [%]	Corr [%]	Acc [%]
5	1	41,4	-36,9	38,7	-50,5
	2	50,5	-16,2	46,0	-28,8
	4	56,8	7,2	51,4	-19,8
	8	65,8	36,9	56,8	-1,8
10	1	42,3	-27,9	42,3	-40,5
	2	53,2	-6,3	46,0	-33,3
	4	64,0	21,6	50,5	-14,4
	8	**71,2**	**48,7**	**58,6**	**1,8**

Tabulka 3: Výsledky testování izolovaných slov s 2. typem gramatiky

Testovací sada obsahuje nahrávky s izolovanými slovy, mezi kterými je vždy krátké ticho (v řádu desetin vteřiny). Při ruční kontrole výsledných souborů, které obsahovaly výpisy rozpoznaných slov (i ruchů a ticha), se zjistilo, že ticho mezi slovy je správně rozpoznáno pouze v několika málo případech. Pro třetí a čtvrté testování byla tedy gramatika upravena tak, že po slově pokaždé následovalo ticho.

V Tabulce 4 jsou uvedeny výsledky pro třetí typ gramatiky (označené G3), která povoluje jakékoliv slovo ze slovníku následované tichem. V tomto testování nebyly uvažovány ruchy. Míra úspěšnosti Corr se oproti předchozímu testování s gramatikou G2 zvedla u trénovací sady Prec Train v průměru o 9,3 %. U sady Approx Train se zvedla míra úspěšnosti Corr o 5,4 %. U míry úspěšnosti Acc se ale úspěšnost zvedla ještě mnohem více. Pro trénovací sadu Prec Train to bylo v průměru o 55 % a pro sadu Approx Train o 61,5 %. Je tedy vidět, že je vhodné mít gramatiku nastavenou přímo na danou aplikaci.

počet iterací	počet mixtur	Prec Train		Approx Train	
		Corr [%]	Acc [%]	Corr [%]	Acc [%]
5	1	47,8	31,5	37,8	17,1
	2	60,4	51,4	45,1	30,6
	4	68,5	60,4	58,6	45,1
	8	76,6	73,0	63,1	49,5
10	1	51,4	36,9	40,5	27,0
	2	61,3	53,2	51,4	37,8
	4	73,0	68,5	58,6	44,1
	8	**81,1**	**80,2**	**61,3**	**44,1**

Tabulka 4: Výsledky testování izolovaných slov s 3. typem gramatiky

Čtvrté testování, které zkoumalo výsledky pro různé počty iterací, mixtur výstupních funkcí a kvalitu fonetických přepisů, přidávalo do gramatiky výše zmíněné ruchy zároveň s vynucenou mezerou po rozpoznaném slově. Tato gramatika byla označena G4. Výsledky jsou uvedeny v Tabulce 5. V průměru se úspěšnost pro obě dvě sady o 2,5 % zlepšila.

počet iterací	počet mixtur	Prec Train		Approx Train	
		Corr [%]	Acc [%]	Corr [%]	Acc [%]
5	1	46,9	35,1	42,3	25,2
	2	62,2	55,0	51,4	39,6
	4	73,9	69,4	60,4	46,9
	8	78,4	74,8	63,1	45,1
10	1	48,7	39,6	44,1	32,4
	2	64,0	57,7	50,5	38,7
	4	74,8	73,0	59,5	46,0
	8	**81,1**	**80,2**	**62,2**	**43,2**

Tabulka 5: Výsledky testování izolovaných slov s 4. typem gramatiky

Při posledním testu byl použit model natrénovaný v Laboratoři počítačového zpracování řeči. Tento model byl natrénován na trénovací sadě obsahující několik hodin záznamů od velkého počtu mluvčích. Modely monofonů byly 9krát iterovány a mají 96 mixtur. Z Tabulky 6 je vidět, že výsledky s poskytnutým modelem jsou výrazně lepší. V případě G1, která povolovala pouze jednotlivá slova ze slovníku a mezeru, dosahoval model Prec Train s 10 iteracemi a 8 mixturami výstupní funkce úspěšnost 45 %. Model poskytnutý Laboratoří počítačového zpracování řeči dosahuje pro tento systém rozpoznávání až 80 % (pro míru úspěšnosti Acc). U gramatiky G2 je rozdíl v míře úspěšnosti Acc mezi nejlepším testovaným modelem Prec Train a monofonovým modelem poskytnutým

Laboratoří počítačového zpracování řeči přibližně 30 %. U dalších dvou použitých gramatik G3 a G4 již nejsou rozdíly tak výrazné, přesto se pohybují v řádech několika procent.

Gramatika	Corr [%]	Acc [%]
G1	86,5	80,2
G2	84,7	78,4
G3	86,5	85,6
G4	84,7	83,8

Tabulka 6: Výsledky testování izolovaných slov s modelem poskytnutým laboratoří SpeechLab

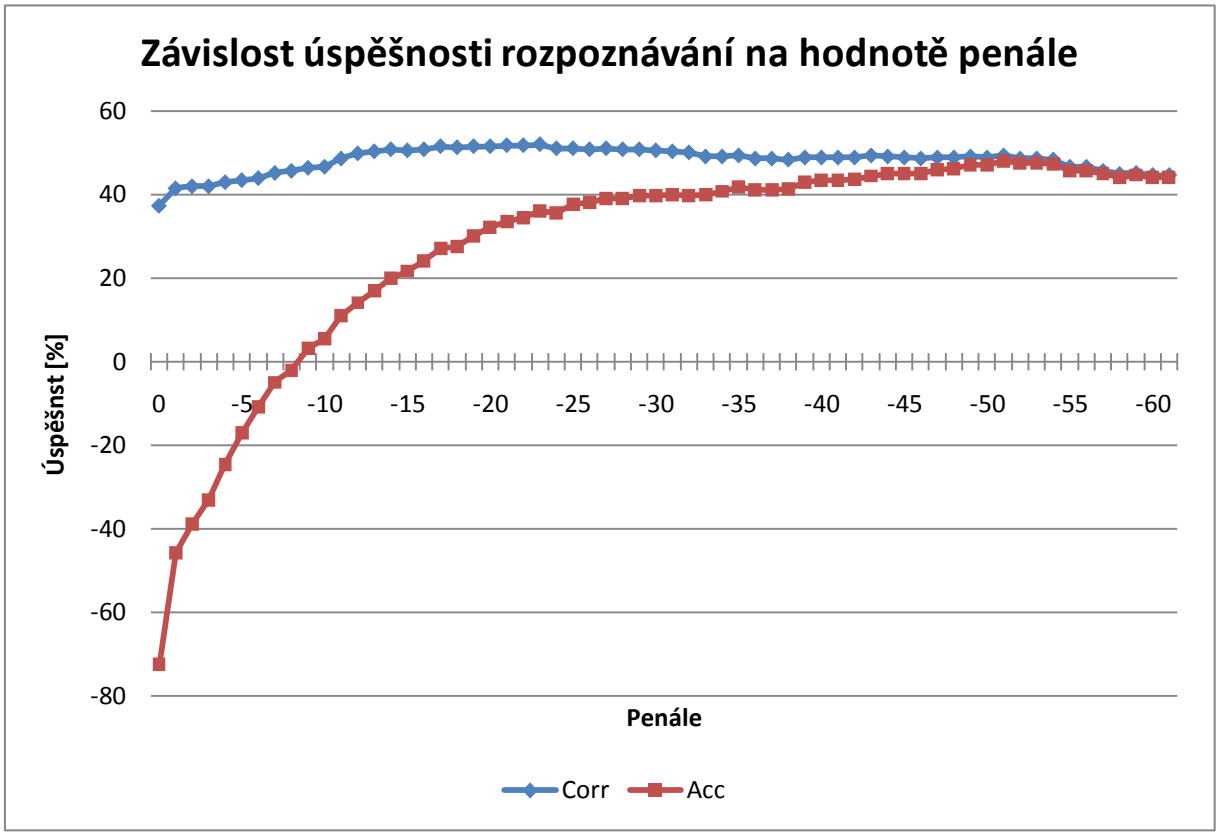
Nahrávky pro testování použité v této kapitole, které obsahovaly izolovaná slova, nebylo možné použít pro otestování vhodné hodnoty penále. Izolovaná slova byla pro toto testování zvolena velmi podobná. Jejich délky si tedy přibližně odpovídaly, takže se pro ně nedalo otestovat vhodné nastavení penále.

Pro testování hodnoty penále byly využity poznatky z předchozího testování. Podle výsledků byl vybrán nejlepší model – sada Prec Train, 10 iterací, 8 mixtur výstupní pravděpodobnostní funkce. Jako testovací nahrávky již nebyly použity nahrávky obsahující izolovaná slova, ale nahrávky se spojitou řečí, takže nemohla být použita nejlepší gramatika z předchozích testů (nejlepší gramatika G4 požadovala za každým rozpoznaným slovem ticho). Byla proto vytvořena nová gramatika, která dovolovala jakékoliv slovo ze slovníku a všechny ruchy popsané u předchozích testů. Slovník se skládal ze slov, která byla obsažena v testovacích nahrávkách a z některých jejich výslovnostních variant. Celkem slovník obsahuje 1264 slov.

Pro testování vhodné hodnoty penále byly použity testovací nahrávky z MultiMic Database. Vzhledem k výpočetní náročnosti byla nejprve použita podmnožina této sady, konkrétně deset různých nahrávek (celkem obsahovaly 409 slov). Po zjištění nejlepší hodnoty penále bylo zvoleno několik hodnot jí blízkých a pro ty bylo testováno s celou sadou testovacích nahrávek (celkem 4121 slov – vybraná podmnožina tedy obsahovala přibližně desetinu z celkového počtu slov), aby byla zjištěná hodnota ověřena pro všechna testovací data.

V grafu na Obrázku 12 je vidět závislost úspěšnosti rozpoznávání na hodnotě penále. Míra úspěšnosti Corr je v grafu zobrazena modře, míra úspěšnosti Acc červeně. Je vidět, že míra úspěšnosti Corr se od penále s hodnotou -10 drží kolem 50% úspěšnosti. Až kolem

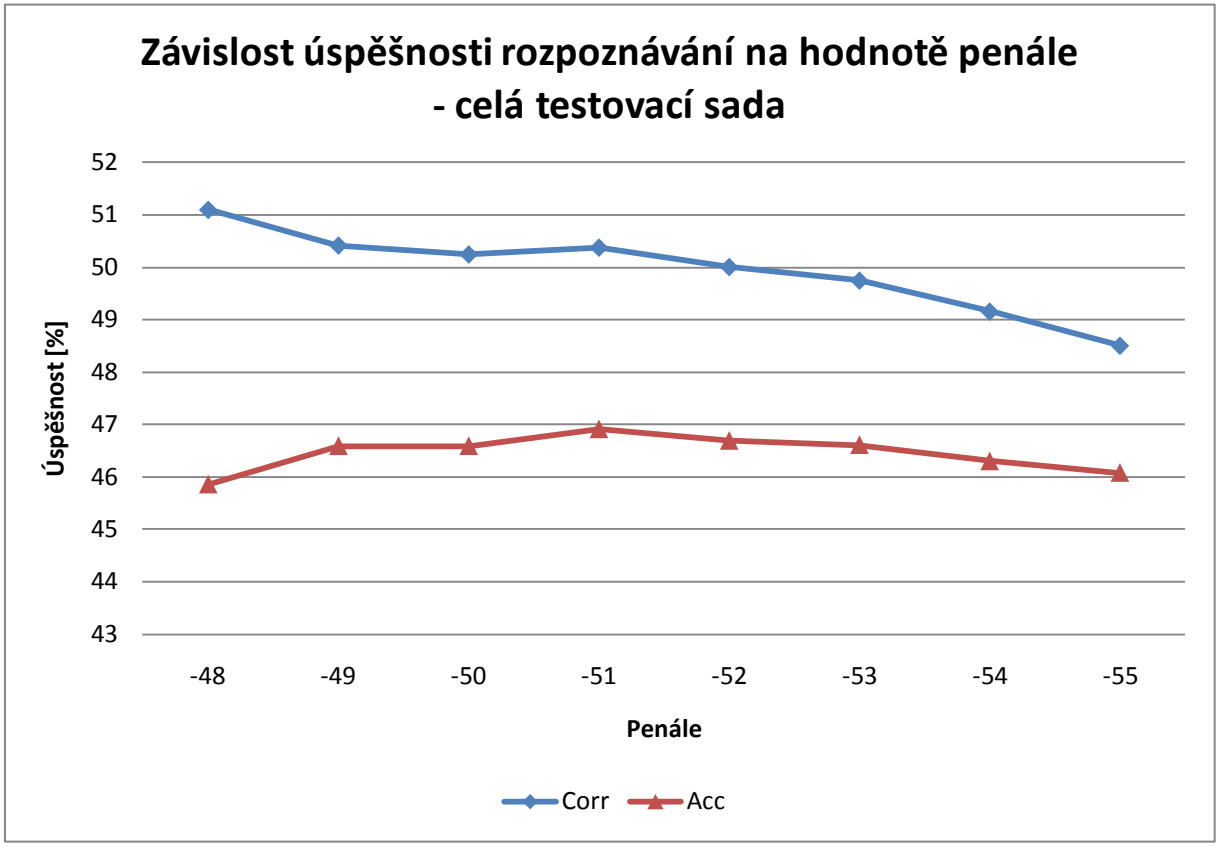
hodnoty penále -55 klesá opět k úspěšnosti rovné 45 %. Míra úspěšnosti Acc má pro penále 0 záporný výsledek, rovný -72 %. Se snižující se hodnotou penále se nejprve velmi rychle zvyšuje hodnota úspěšnosti, postupně se ale úspěšnost zvyšuje pomaleji. Pro hodnoty penále nižší než -55 začíná míra úspěšnosti Acc opět klesat. Nejvyšší hodnota úspěšnosti rozpoznávání byla pro penále -51. Míra úspěšnosti Acc dosahovala 48,2 % a míra úspěšnosti Corr byla rovna 49,4 %.



Obrázek 12: Závislost úspěšnosti rozpoznávání na hodnotě penále - pro podmnožinu testovací sady

V grafu na následujícím Obrázku 13 bylo provedeno stejné testování jako v předchozím případě, ale testovalo se se všemi testovacími nahrávkami a testování se omezilo na penále blízké hodnotě -51, které v předchozím testování dopadlo nejlépe. Míra úspěšnosti Corr (v grafu vyznačena modrou barvou) má se snižujícím se penále klesající tendenci, stejně jako v předchozím případě. Dosahuje v průměru o jedno procento lepší úspěšnosti než u vybrané podmnožiny testovacích nahrávek. Míra úspěšnosti Acc má při hodnotě penále -51 nejvyšší úspěšnost, rovnou 46,9 %, což je o něco nižší než v předchozím testování, kde dosahovala úspěšnost v tomto bodě 48,2 %. Křivky mají stejné tendence jako při testování na podmnožině testovací sady. Nejlepší úspěšnost byla dosažena opět pro penále

-51. Při změně rozpoznávacího systému může dojít i k posunu nejlepší hodnoty penále, byla proto pro další testování zvolena hodnota -50.



Obrázek 13: Závislost úspěšnosti rozpoznávání na hodnotě penále - pro celou testovací sadu

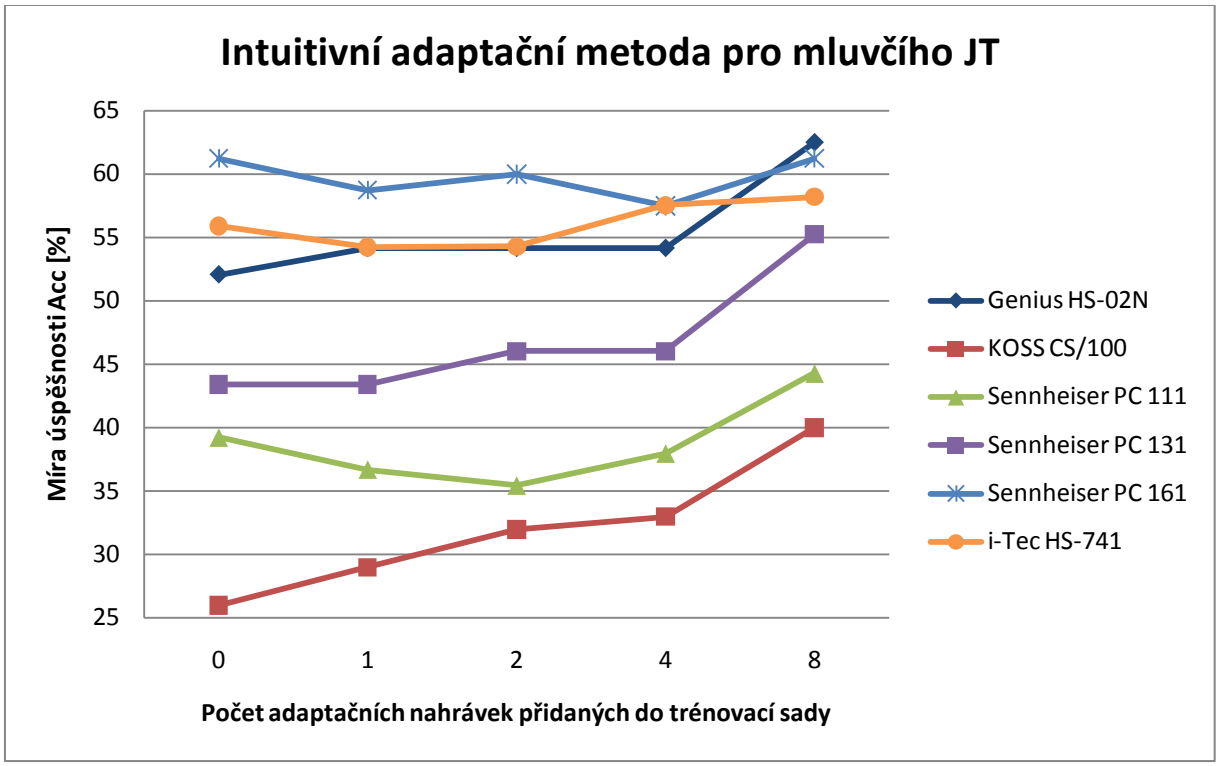
6.1.3 Pokročilejší experimenty

Další testování bylo již zaměřeno na adaptaci systému pro rozpoznávání řeči na konkrétního mluvčího a mikrofon. Pro tuto úlohu byla využita hlavní databáze nahrávek MultiMic Database a to jak testovací nahrávky, tak i adaptační. Pro testování se využily výsledky z předchozí kapitoly.

První testovaná metoda adaptace byla intuitivní metoda, která ilustrovala základní principy adaptace. K sadě trénovacích nahrávek byla přidána adaptační nahrávka získaná od jednoho konkrétního mluvčího a z jednoho mikrofonu. Tím se zajistilo, že modely natrénované z trénovací sady (sada Prec Train, která obsahuje nahrávky od jiných mluvčí než testovací sada MultiMic Database) se přiblížily k danému mluvčímu a přenosovému kanálu. Trénovací sada ovšem obsahuje několikanásobně větší objem dat než jedna adaptační

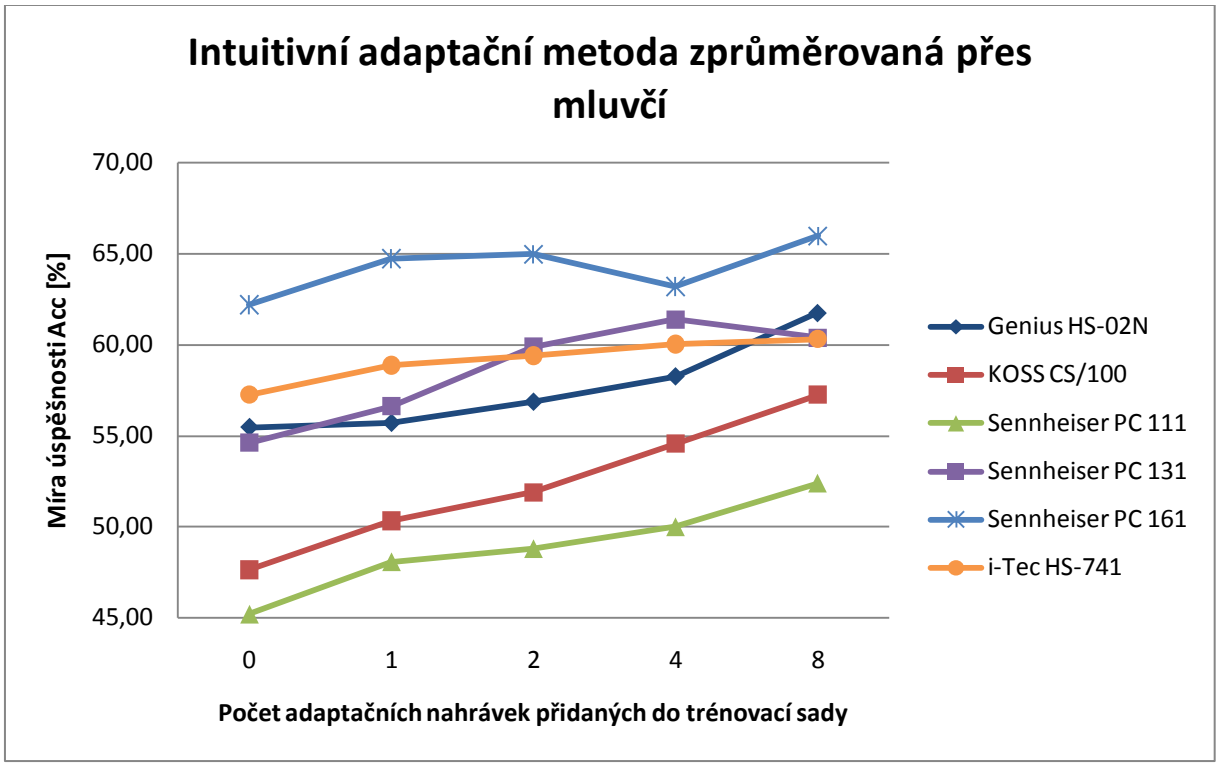
nahrávka. Bylo tedy také otestováno to, jaký bude mít vliv, když zvýšíme váhu adaptačních dat. Zvýšení váhy bylo dosaženo tím, že adaptační data byla k trénovací sadě přidána vícekrát. Konkrétně byla adaptační nahrávka přidána 1krát, 2krát, 4krát a 8krát. Testování proběhlo na nahrávce od stejného mluvčího a na stejném přenosovém kanále jako adaptační nahrávka. Mohlo se tak posoudit, jaká byla změna úspěšnosti rozpoznávání s adaptovaným modelem oproti modelu, který adaptován nebyl. Aby byly výsledky testování dostatečně prokazatelné, byly tyto testy provedeny se všemi nahrávkami z použité databáze MultiMic Database.

V grafu na Obrázku 14 jsou vidět výsledky testování intuitivní metody adaptace pro jednoho konkrétního mluvčího (tabulky s výsledky pro všechny mluvčí jsou uvedeny v Příloze D). Jedna řada v grafu reprezentuje jednu nahrávku od daného mluvčího a mikrofon, který je uveden v legendě. Pouze řada s názvem i-Tec HS-741 reprezentuje průměrné hodnoty 5 různých nahrávek zvoleného mluvčího. Úspěšnost (konkrétně je v grafu zobrazena míra úspěšnosti Acc) se s postupným přidáváním adaptačních nahrávek k trénovací sadě v průměru zvyšovala. U konkrétních nahrávek ale občas došlo i k poklesu úspěšnosti. Například nahrávka z mikrofonu Sennheiser PC 161 měla po adaptaci modelů 1 adaptační nahrávkou úspěšnost nižší o 3 %. Při rozpoznávání s modely monofonů adaptovanými 8 adaptačními nahrávkami měla stejnou úspěšnost jako při rozpoznávání s neadaptovanými modely. Naopak nahrávka nahraná mikrofonem KOSS CS/100 byla rozpoznána při každé přidané adaptační nahrávce o několik procent lépe. Při rozpoznávání neadaptovaným modelem dosahovala úspěšnost rozpoznávání 26 %, při rozpoznávání s modelem adaptovaným 8 adaptačními nahrávkami, dosahovala úspěšnost rozpoznávání 40 %. Úspěšnost tedy stoupla o 14 %.



Obrázek 14: Graf výsledků intuitivní adaptační metody pro jednoho konkrétního mluvčího JT

Graf na Obrázku 15 zobrazuje průměrné výsledky pro intuitivní adaptační metodu. Výsledky jsou zprůměrované přes všechny mluvčí a nahrávky z mikrofonu i-Tec HS-741 jsou navíc ještě zprůměrované přes všech pět nahrávek od jednoho mluvčího. Z grafu je vidět, že v průměru se po přidání jedné adaptační nahrávky k trénovací sadě úspěšnost zlepší v řádu jednotek procent. Při přidání více adaptačních dat se úspěšnost průměrně zlepšila o něco více. V tomto případě obsahují adaptační data 63 slov a v průměru trvají 30 vteřin. Proto se průměrné zlepšení úspěšnosti pohybuje v řádech jednotek procent.



Obrázek 15: Graf výsledků intuitivní adaptační metody, výsledky jsou uvedeny v míře úspěšnosti Acc zprůměrované přes všechny mluvčí

6.1.4 Shrnutí úvodních experimentů

Tato kapitola sloužila k seznámení s prostředím HTK v souladu s bodem 4 u cílů práce. Bylo otestováno rozpoznávání izolovaných slov, u kterých se zjišťovala závislost výsledné úspěšnosti rozpoznávání na počtu iterací modelu, počtu mixtur výstupních funkcí a především na kvalitě fonetických přepisů trénovacích nahrávek. Experimentálně bylo zjištěno, že v případě izolovaných slov velmi závisí na použitém jazykovém modelu. Pro toto testování nejlépe vyšla pevná gramatika připouštějící jakékoliv slovo ze slovníku a za ním vždy následovalo ticho. Pro trénovací sadu nahrávek Prec Train dosahovala úspěšnost až 80% v míře Acc.

Dále byla otestovaná hodnota penále. Nejprve bylo penále otestováno pro podmnožinu testovací sady MultiMic Database a poté bylo otestováno pro celou sadu, už ale pro menší počet hodnot vybraných předchozím testováním. Jako nejvhodnější hodnota penále byla experimentálně zjištěna hodnota -50.

Mezi úvodní experimenty patřila i jedna metoda adaptace, která daný model přiblížila ke konkrétnímu mluvčímu a k mikrofonu. I přes to, že tato metoda patří k úvodním testům, dosáhla poměrně dobrých výsledků, které jsou vidět v grafech v kapitole 6.1.3 Pokročilejší experimenty a také v tabulkách s konkrétními hodnotami, které jsou v Příloze D.

6.2 Experimenty s klíčovými metodami adaptace

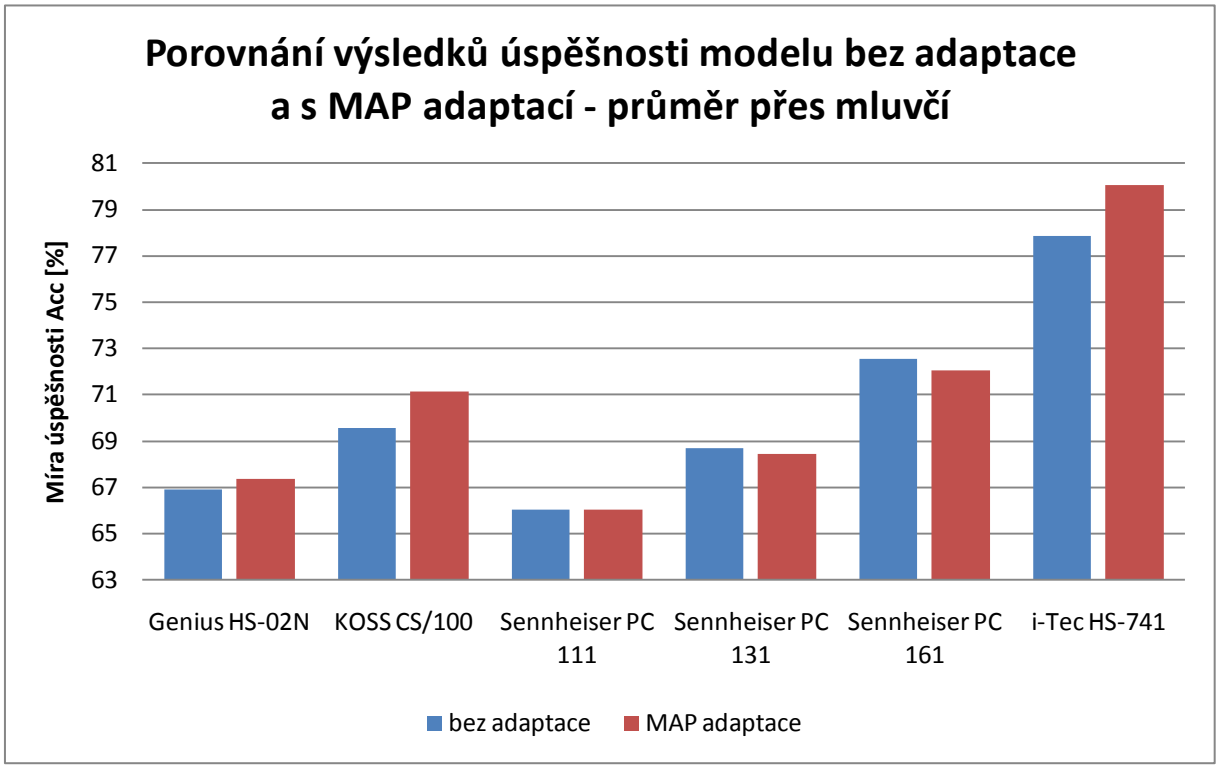
Tato kapitola popisuje dvoje testování různých adaptačních metod. První testování využilo natrénovaných modelů monofonů z Laboratoře počítačového zpracování řeči. Tyto modely byly adaptovány různými metodami adaptace v prostředí HTK a následně byly použity pro rozpoznávání systémem vytvořeným v předchozí kapitole, taktéž v prostředí HTK. Druhé testování využilo nejnovějších modelů trifonů natrénovaných v Laboratoři počítačového zpracování řeči, které byly opět adaptovány vybranými metodami v prostředí HTK. Modely byly použity pro rozpoznávání řeči v programu používaném v reálných aplikacích.

6.2.1 Testování metod adaptace pro modely monofonů

Pro testování adaptačních metod v prostředí HTK byl použit rozpoznávací systém stejný jako v Kapitole 6.1.3 Pokročilejší experimenty. Opět se testovaly testovací nahrávky z MultiMic Database, slovník i gramatika zůstaly stejné. V tomto testování byly využity modely monofonů natrénované v Laboratoři počítačového zpracování řeči. Pro testování byla nejprve vybrána metoda MAP. Adaptační váha pro adaptaci vektorů středních hodnot všech mixtur byla zvolena $\tau_{ik} = 15$. Tabulky výsledků pro všechny testované metody jsou uvedeny v Příloze E.

V grafu na Obrázku 16 jsou uvedeny výsledky testování úspěšnosti rozpoznávání s neadaptovaným modelem (modré sloupce) a s modelem adaptovaným metodou MAP (červené sloupce) zprůměrované přes všechny mluvčí. Z výsledků v grafu je vidět, že metoda MAP není příliš vhodná pro adaptaci s menším objemem adaptačních dat, jako je tomu v tomto případě. Například u nahrávky z mikrofonu KOSS CS/100 se úspěšnost rozpoznávání zlepšila o 1,5 %. Ale naopak pro mikrofon značky Sennheiser PC 161 se úspěšnost

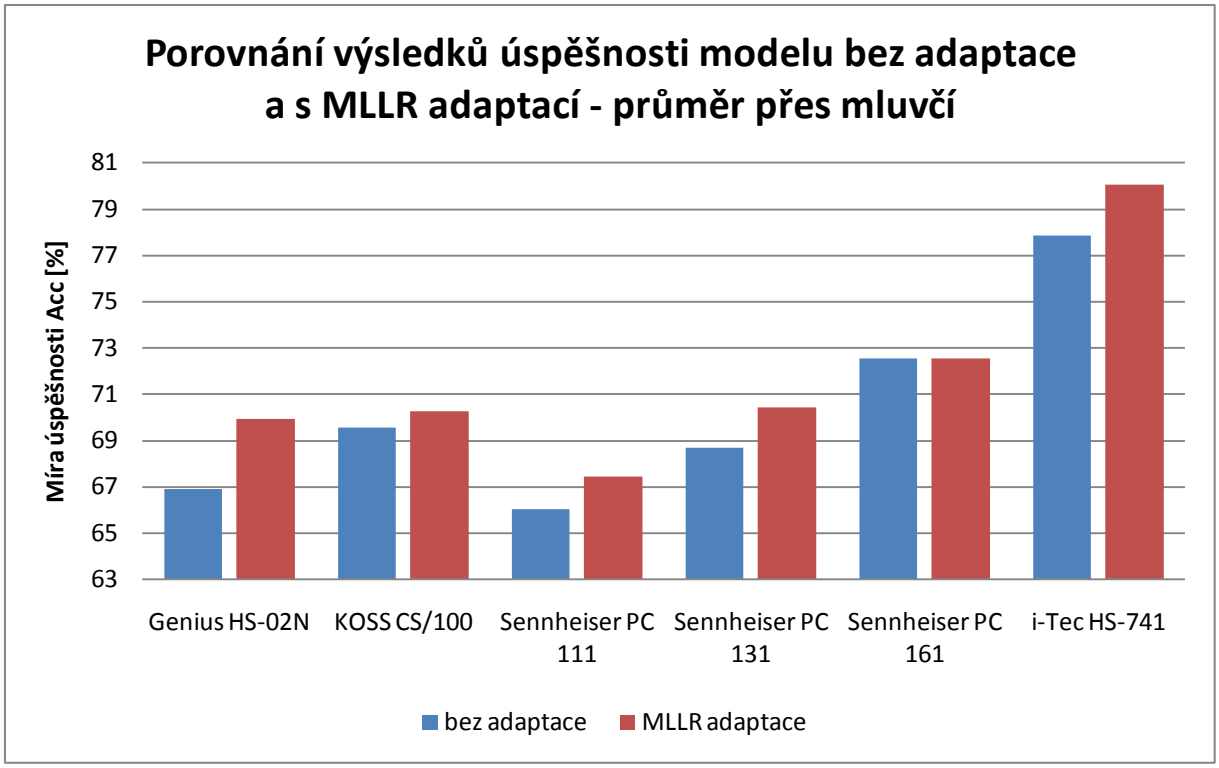
rozpoznávání o 0,5 % zhoršila. V průměru se míra úspěšnosti zlepšila po adaptaci metodou MAP o 0,6 %. Jak již bylo zmíněno, adaptační nahrávky z vytvořené MultiMic Database jsou v průměru 30 vteřin dlouhé. Metoda MAP sice s rostoucím množstvím adaptačních dat konverguje k teoreticky nejlepšímu SD modelu, ale v případě, že je k dispozici menší množství dat (v řádu jednotek minut), tak se neadaptují modely všech fonémů, popřípadě se adaptují nedostatečně.



Obrázek 16: Graf znázorňující úspěšnosti rozpoznávání s neadaptovaným modelem (modré sloupce) a s modelem adaptovaným metodou MAP (červené sloupce) zprůměrované přes všechny mluvčí

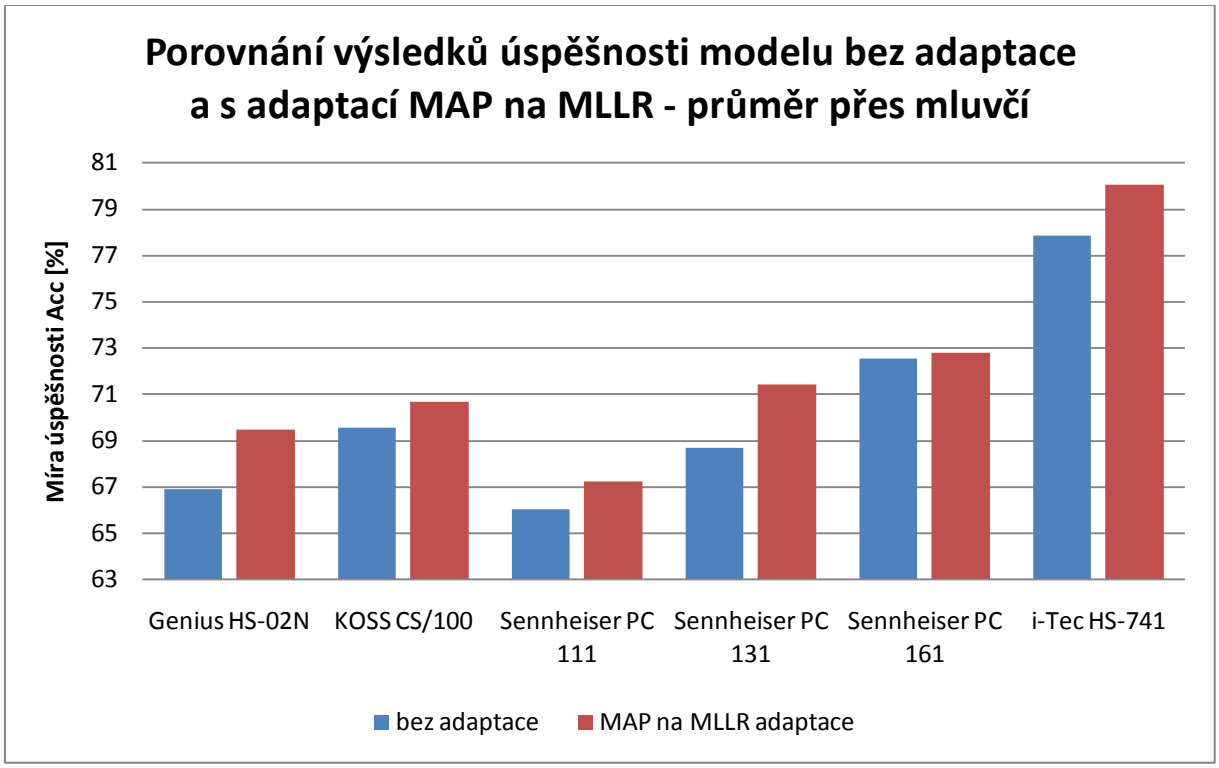
Další testovanou metodou byla metoda MLLR. V prostředí HTK byl pro model, který se metodou adaptoval, vytvořen regresní strom, podle kterého byly adaptovány jednotlivé skupiny fonémů. V grafu na Obrázku 17 jsou vidět výsledky pro rozpoznávání s modelem adaptovaným metodou MLLR (červené sloupce) a pro porovnání i výsledky rozpoznávání s původním, neadaptovaným modelem (modré sloupce) zprůměrované přes všechny mluvčí. Až na mikrofon značky Sennheiser PC 161 se v průměru zvedla míra úspěšnosti Acc u všech mikrofonů. Největší rozdíl v úspěšnosti nastal u mikrofonu značky Genius HS-02N, pro který se úspěšnost zvedla o 3 %. V průměru se úspěšnost zlepšila o necelé 2 %. Vzhledem k rozsahu použitých adaptačních dat je zlepšení v řádu jednotek procent dobré. Stejně jako u

předchozí metody MAP byly potvrzeny teoretické předpoklady. U metody MLLR je výhodou, že jedna transformační matice může být vypočtena pro více Gaussových komponent, které jsou si podobné (tuto podobnost hledá právě vytvořený regresní strom). Je tedy možné použít menší množství adaptačních dat.



Obrázek 17: Graf znázorňující úspěšnosti rozpoznávání s neadaptovaným modelem (modré sloupce) a s modelem adaptovaným metodou MLLR (červené sloupce) zprůměrované přes všechny mluvčí

Pro třetí testování byla použita metoda, která kombinovala předchozí dvě metody. Parametry modelů byly nejprve transformovány metodou MLLR a poté byly estimovány vektory středních hodnot metodou MAP. Výsledky tohoto testování jsou obdobné jako výsledky z testování metody MLLR, ale mají v průměru o 0,5 % lepší úspěšnost. Graf výsledků zprůměrovaný přes všechny mluvčí je vidět na Obrázku 18. Pro porovnání jsou v grafu zobrazeny i výsledky neadaptovaného modelu.



Obrázek 18: Graf znázorňující úspěšnosti rozpoznávání s neadaptovaným modelem (modré sloupce) a s modelem adaptovaným metodou MAP na MLLR (červené sloupce) zprůměrované přes všechny mluvčí

6.2.2 Vyhodnocení metod adaptace pro modely monofonů

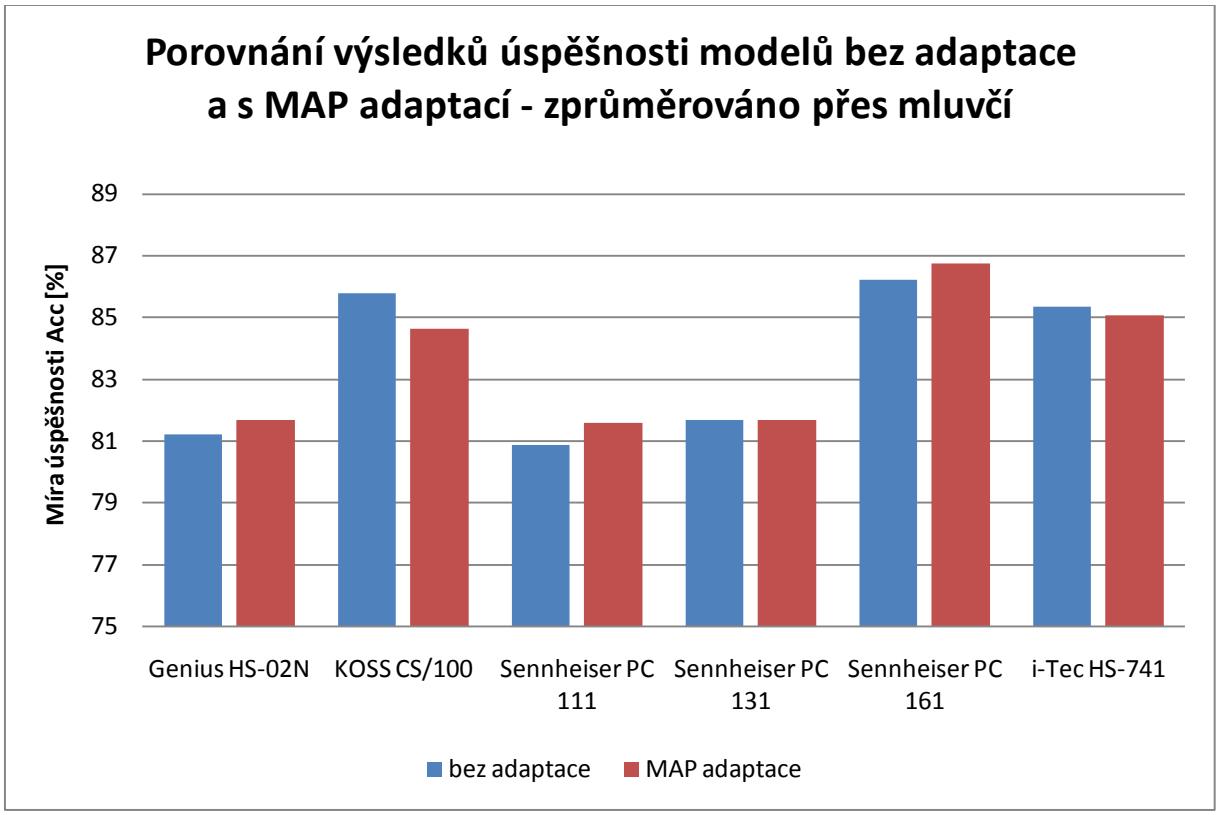
V této kapitole bylo otestováno několik metod adaptace modelů monofonů v prostředí HTK. Z výsledků je vidět, že metoda MAP nedosahovala při použitých adaptačních nahrávkách velkého zlepšení úspěšnosti. Naopak metoda MLLR zvedla průměrnou úspěšnost o necelá 2 %, což je při použitých adaptačních nahrávkách znatelný výsledek. Třetí testovaná metoda, která vycházela v předchozích dvou metod, dosahovala výsledků přibližně stejných jako metoda MLLR.

6.2.3 Testování metod adaptace pro modely trifonů

Všechny tři testované adaptace popsané v předchozí kapitole adaptovaly modely monofonů. V zadání diplomové práce bylo i otestování adaptace modelů trifonů. Modely trifonů je potřeba natrénovat na rozsáhlé trénovací sadě, byly proto využity již natrénované modely trifonů, poskytnuté Laboratoří počítačového zpracování řeči. Systém pro rozpoznávání řeči vyvinutý v Laboratoři počítačového zpracování řeči pracuje s jiným typem přepisů trifonů, než systém pro rozpoznávání vytvořený v prostředí HTK a proto klesla úspěšnost rozpoznávání při použití tohoto systému až o 25 %. Testování bylo tedy provedeno pomocí programu na rozpoznávání řeči vyvinutém v Laboratoři počítačového zpracování řeči. Použitý systém pro rozpoznávání obsahuje slovník se 338 679 slovy a pravděpodobnostním jazykovým modelem.

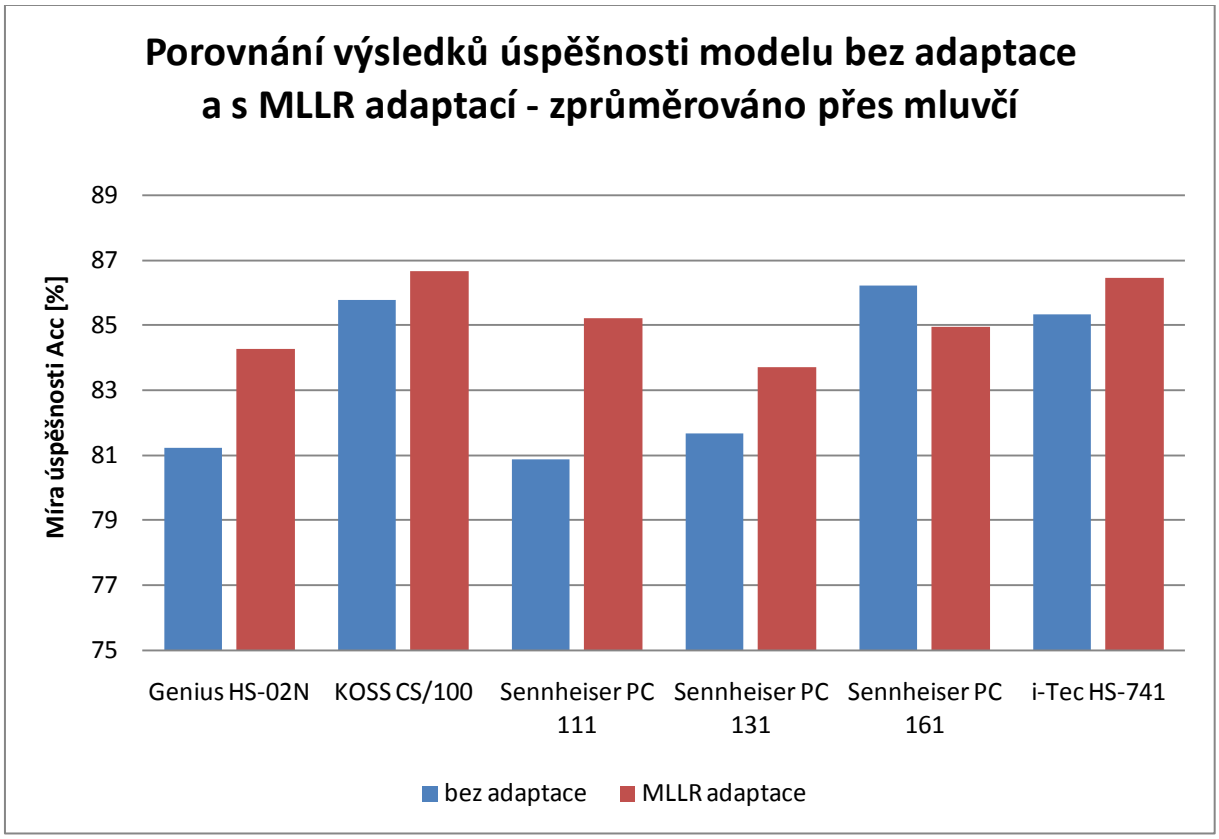
Samotná adaptace modelů byla provedena v prostředí HTK, bylo však potřeba zajistit převod fonetických přepisů adaptačních nahrávek z monofonů na trifony a poté upravit výsledný adaptovaný model tak, aby byl kompatibilní s rozpoznávačem řeči. Převod fonetických přepisů zajistil z části nástroj v HTK a z části předem připravený skript napsaný v programovacím jazyce Perl. Převod výsledného modelu byl proveden pomocí dalšího skriptu napsaného v programovacím jazyce Perl.

V grafu na Obrázku 19 jsou vidět výsledky testování adaptace modelů metodou MAP (červené sloupce) zprůměrované přes všechny mluvčí. Konkrétní výsledky pro všechny mluvčí a všechny metody jsou uvedeny v Příloze F. V grafu jsou pro porovnání zobrazeny i výsledky neadaptovaného modelu (modré sloupce). I při použití reálného systému pro rozpoznávání řeči nedosahuje metoda MAP při použitých adaptačních datech výrazného zlepšení. Pro mikrofon značky KOSS CS/100 dokonce došlo ke zhoršení výsledné úspěšnosti o 1 %. V celkovém průměru přes všechny nahrávky došlo ke zhoršení o několik desetin procenta.



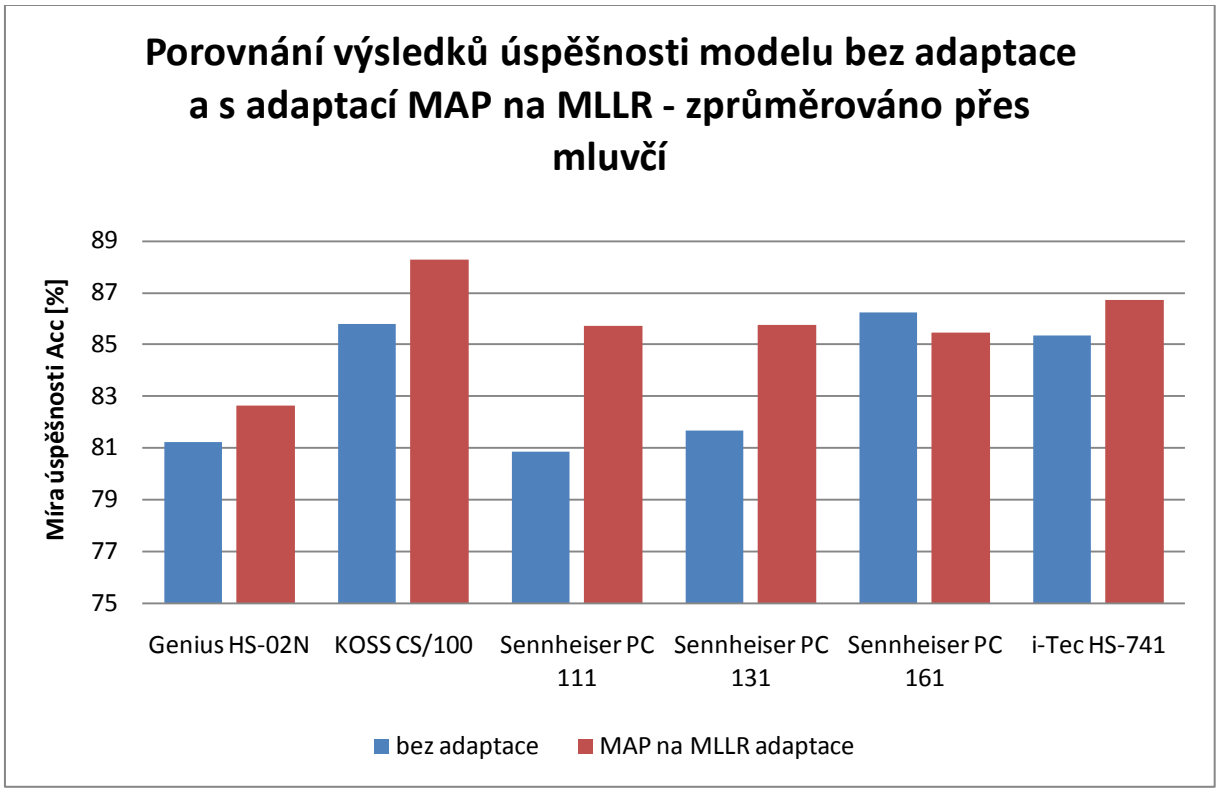
Obrázek 19: Graf znázorňující úspěšnosti rozpoznávání s neadaptovaným modelem (modré sloupce) a modelem adaptovaným metodou MAP (červené sloupce) zprůměrované přes všechny mluvčí

Na následujícím grafu na Obrázku 20 jsou vidět výsledky rozpoznávání pro modely adaptované metodou MLLR. U tohoto testování již došlo ke znatelnému zlepšení výsledků. Pouze u mikrofonu značky Sennheiser PC 161 se míra úspěšnosti Acc zhoršila přibližně o 1,5 %. Je vidět, že adaptace metodou MLLR nejenom zvýšila úspěšnost rozpoznávání, ale také potlačila rozdíly mezi úspěšnostmi nahrávek z různých mikrofonů a od různých mluvčích. Zatímco rozdíly úspěšností nahrávek z různých mikrofonů rozpoznávaných neadaptovanými modely přesahují 5,3 %, rozdíl úspěšností u nahrávek rozpoznaných modely adaptovanými metodou MLLR činí necelá 3 %.



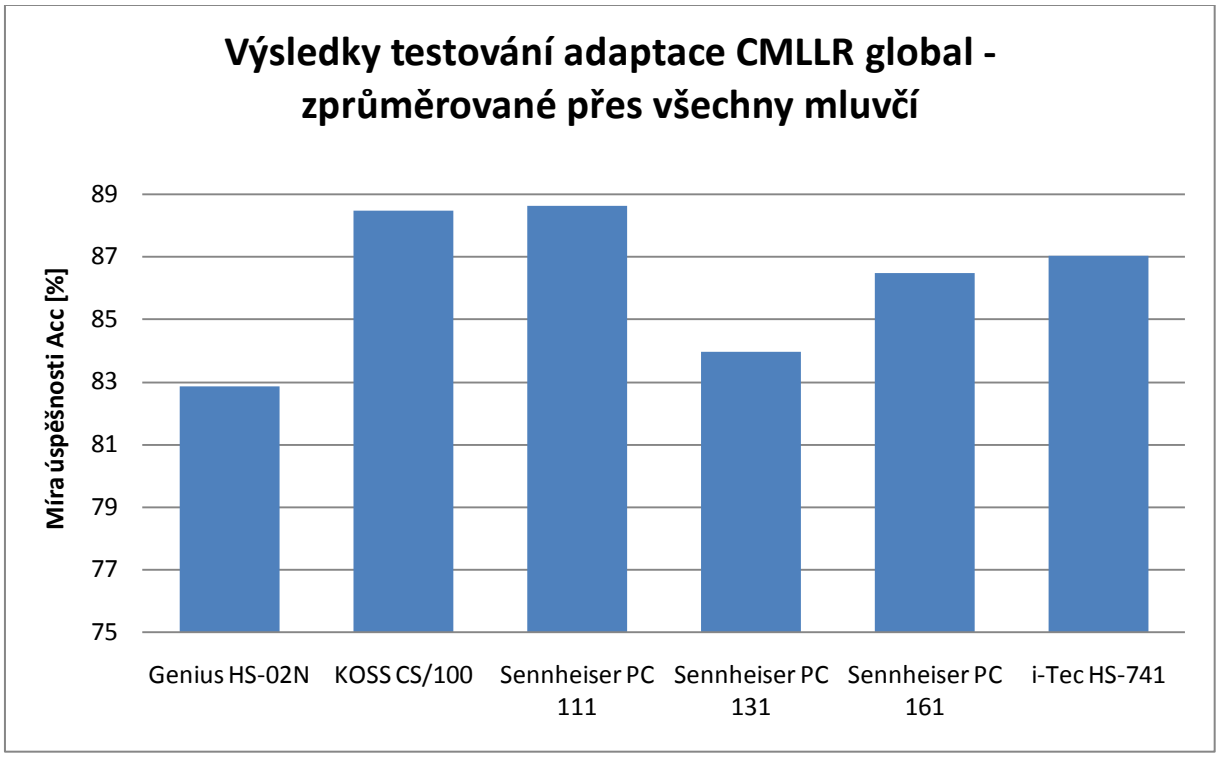
Obrázek 20: Graf znázorňující úspěšnosti rozpoznávání s neadaptovaným modelem (modré sloupce) a s modelem adaptovaným metodou MLLR (červené sloupce) zprůměrované přes všechny mluvčí

V grafu na Obrázku 21 jsou zobrazeny výsledky testování pro kombinaci metod MLLR a MAP. Výsledky z tohoto testování jsou o něco vyšší než výsledky testování samotné metody MLLR.



Obrázek 21: Graf znázorňující úspěšnosti rozpoznávání s neadaptovaným modelem (modré sloupce) a s modelem adaptovaným metodou MAP na MLLR (červené sloupce) zprůměrované přes všechny mluvčí

Pro testování na reálném systému pro rozpoznávání řeči byla zvolena ještě metoda CMLLR, která neadaptovala model, ale transformovala příznakové vektory testovacích nahrávek. Metoda umožňuje transformovat všechny příznakové vektory stejnou transformací nebo, stejně jako metoda MLLR, transformuje příznakové vektory podobných fonémů jednou transformací. Vzhledem k malému objemu adaptačních dat byla zvolena globální transformace, takže se všechny příznakové vektory transformovaly stejnou transformací. Pro testování byly použity odlišné modely fonémů natrénované ze sady nahrávek, kterým se po parametrizaci transformovaly příznakové vektory (SAT). Není proto možné porovnávat úspěšnosti s neadaptovaným modelem. Průměrné výsledky z této metody jsou uvedeny v grafu na Obrázku 22. Rozdíly mezi jednotlivými úspěšnostmi z různých mikrofonů jsou opět až 5 %.



Obrázek 22: Graf zobrazující úspěšnosti rozpoznávání řeči s použitou metodou adaptace CMLLR

6.2.4 Vyhodnocení metod adaptace pro modely trifonů

Testování metod adaptace bylo provedeno na profesionálním systému rozpoznávání řeči poskytnutým Laboratoří počítačového zpracování řeči. Výsledné úspěšnosti rozpoznávání nedosahovaly předpokládaných hodnot, to bylo způsobeno především výběrem vět. Testování různých metod adaptace dopadlo obdobně jako při testování monofonových modelů v systému rozpoznávání vytvořeném v prostředí HTK. Metoda MAP nepřinesla výrazné zlepšení úspěšnosti. Naopak metoda MLLR, stejně jako u předchozích testů, dosáhla znatelného zlepšení úspěšností. Byla také otestována metoda CMLLR, která se ovšem nedá porovnat s výsledky z neadaptovaných modelů, protože používá speciálně navržený model fonémů.

7. Závěr

V rámci této diplomové práce byla vytvořena databáze nahrávek, které se dají objektivně využít k porovnávání různých mikrofonů. Tato databáze obsahuje více než jednu hodinu nahrávek od šesti různých mluvčích a z šesti různých mikrofonů. Je navíc specifická tím, že nahrávky byly pořízeny vždy paralelně pro dvojici mikrofonů, z nichž jeden byl referenční. Díky této databázi se ukázalo, že mikrofon použitý při nahrávání má nezanedbatelný vliv na úspěšnost rozpoznávání. Při testování jednotlivých nahrávek s modely monofonů byly rozdíly mezi nahrávkami z různých mikrofonů skoro až 20 % (v míře úspěšnosti Acc) pro konkrétní nahrávky a necelých 6 % z výsledků zprůměrovaných přes všechny mluvčí. Při rozpoznávání s modely trifonů se celková úspěšnost rozpoznávání výrazně zvedla a rozdíly mezi mikrofony byly částečně potlačeny. Maximální rozdíl mezi dvěma různými mikrofony pro konkrétního mluvčího byl přibližně 12 %. Oproti původním necelým 20 % je tento rozdíl znatelně lepší. Největší rozdíl výsledků zprůměrovaných přes mluvčí byl 3 %.

Byl vytvořen a úspěšně otestován systém pro rozpoznávání izolovaných slov v prostředí HTK a bylo natrénováno několik různých modelů. Testování se zaměřilo na různé parametry modelu, jako například na kvalitu fonetických přepisů, počet mixtur výstupní funkce a podobně. Testováním bylo zjištěno, že úspěšnost rozpoznávání je závislá na fonetických přepisech a že je třeba, aby tyto přepisy byly velmi pečlivě vytvořeny.

V prostředí HTK byl vytvořen systém pro rozpoznávání spojité řeči, na kterém bylo otestováno několik adaptačních metod. Metoda MAP se neukázala jako příliš vhodná pro adaptaci v situaci, kdy nejsou adaptační data příliš rozsáhlá. Podle teoretických předpokladů by bylo vhodné mít adaptační data v rozsahu alespoň deseti minut. Adaptační data použitá z vytvořené databáze obsahovala přibližně 30 vteřin promluvy. I přes takto malé množství dat prokázala další testovaná metoda, metoda MLLR, viditelné zlepšení úspěšnosti rozpoznávání. Měřeno mírou Acc zlepšila všechny nahrávky v průměru o necelá 2 %.

Testování adaptačních metod proběhlo také na profesionálním systému pro rozpoznávání řeči, který byl vyvinut v Laboratoři počítačového zpracování řeči a využívá se v komerčních aplikacích. Na tomto systému se již pracovalo s modely trifonů, bylo tedy nejprve potřeba adaptovat modely trifonů. Také se musela vyřešit kompatibilita adaptovaných modelů s novým rozpoznávačem. Úspěšnost rozpoznávání s neadaptovaným modelem byla

v míře Acc v průměru pro všechny nahrávky o necelých 15 % lepší (oproti rozpoznávání s neadaptovaným monofonovým modelem), v průměru dosahovala 84,3 %. Zlepšení bylo dáno jednak tím, že použitý model byl model trifonový a byl natrénovaný na větší sadě trénovacích nahrávek. Z velké části to bylo dáno také tím, že použitý systém pro rozpoznávání obsahoval nejnovější slovník (obsahující více výslovnostních variant) a pravděpodobnostní jazykový model.

Na zapůjčeném systému pro rozpoznávání se opět otestovaly metody adaptace MAP a MLLR a navíc se ještě ověřila metoda CMLLR. Adaptační i testovací nahrávky byly použity stejné jako v předchozích testech, tedy z vytvořené databáze pořízené z různých mikrofonů. Podle předpokladů nedopadla úspěšnost rozpoznávání po adaptaci metodou MAP o mnoho lépe než v předchozím případě. V průměru přes všechny nahrávky dokonce o pár setin procenta zhoršila výslednou úspěšnost v míře Acc. Adaptace MLLR zlepšila výslednou úspěšnost v průměru o 1,5 %, dosahovala tedy v průměru 85,8 %. Adaptace CMLLR, která transformovala příznakové vektory neznámé nahrávky, měla průměrnou úspěšnost 86,6 %. Tato úspěšnost se nedá přímo porovnávat s ostatními výsledky, protože při této adaptační metodě byly použity modely fonémů, které byly vytvořeny z transformovaných příznakových vektorů.

Všechny cíle diplomové práce byly splněny. Další, navazující práce by mohla otestovat, zda se vliv rozdílnosti mikrofonů potlačí při použití většího množství adaptačních dat. Pro takové testy by bylo potřeba vytvořit novou a rozsáhlejší databázi nahrávek. Zajímavé výsledky by mohly vyjít též z testu závislosti velikosti adaptačních dat na celkové úspěšnosti rozpoznávání.

Seznam použité literatury

[Červa04] Červa, P.: Metody adaptace systému rozpoznávání řeči na konkrétního mluvčího. Liberec, 2004. Diplomová práce na TUL. Vedoucí diplomové práce Jan Nouza.

[Červa07] Červa, P.: Řízená a neřízená adaptace na mluvčího v systémech rozpoznávání řeči. Liberec, 2007. Disertační práce na TUL.

[Hájek94] Hájek, D., Nouza, J.: Real-Time HMM-Based Isolated Word Recognition System. In Abstracts of 4th Czech-German Workshop "Speech Processing", Praha, 1994.

[Nouza97] Nouza, J., Psutka, J., Uhlíř, J.: Phonetic Alphabet for Speech Recognition of Czech. In: Radio Engineering, vol. 6, no. 4, December 1997.

[Nouza00] Nouza, J., Nejedlová, D.: Experiments with Read Speech Recognition in Czech. Proc. Of 10th Czech-German Workshop "Speech Processing", Praha, 2000.

[Nouza05] Nouza, J., Žďánský, J., David, P., Červa, P., Kolorenč, J., Nejedlová, D.: Fully Automated System for Czech Spoken Broadcast Transcription with Very Large (300K+) Lexicon. In: Proc. of Interspeech 2005, Lisboa, Portugal, 2005.

[Nouza09] Nouza, J., Koldovský, Z., Vích, R. (editoři): Řeč a počítač. Sborník článků. Liberec, 2009. ISBN 978-80-7372-548-8

[Young09] Young, S., Evermann, G., Gales, M.: HTK book. Cambridge University, Engineering Department, 2009. [online]. [cit. 3.5.2011]. URL: <http://htk.eng.cam.ac.uk/docs/docs.shtml>

Příloha A – ukázky promluv z vytvořené databáze paralelních nahrávek MultiMic Database, která byla použita pro porovnávání mikrofonů

Adaptační věty (pro všechny adaptační promluvy stejné):

Archeologové poprvé doložili osídlení Velehradu v době Velké Moravy.

V místě baziliky našli keramiku z devátého století.

Kosterní pozůstatky ze třináctého století byly objeveny na velehradské bazilice.

Náš nález dokládá lidské aktivity přímo na místě baziliky ještě před jejím založením, upřesnil význam nálezu Čermák.

Leckde to ovšem není možné dokázat.

Děláme proto funkční odhady.

Věř, že džus má spoustu vitamínů.

Kód tramvaje je neznámý.

Testovací věty (ukázka jedné sady vět, každá nahrávka měla jiné věty):

Je to velmi nebezpečné. Sloupy se totiž lámou hned u chodníku, kde mají kritické místo, takže padají na ulici z veliké výšky a poměrně rychle.

Navíc jsou pod proudem, uvedla Lea Tomková, podle které zkracují životnost sloupů močící psi nebo solení chodníků. To vše totiž urychluje jejich korozi.

Tomková dodává, že se Služby města snaží nejvíce poškozené sloupy měnit. Práce však nejsou dost rychlé. Celkem je v Pardubicích jedenáct tisíc lamp.

Ročně měníme asi pět set sloupů. Navíc stále hledáme ty poškozené. Nyní se třeba chystá výměna lamp na Dubině.

Příloha B – výsledky testování rozdílů úspěšností rozpoznávání mezi různými mikrofony – testování bylo provedeno pomocí rozpoznávače poskytnutého Laboratoří počítačového zpracování řeči s modely monofonů a trifonů

Testování s modely **monofonů**:

	Míra úspěšnosti Acc [%]						
	M - JB	M - JT	M - PZ	Z - MK	Z - PR	Z - VM	Průměr
Genius HS-02N	77,5	62,5	66,2	69,3	54,9	76,8	67,9
i-Tec HS-741	80,0	55,6	63,5	75,3	73,8	78,3	71,1
KOSS CS/100	78,1	56,4	71,5	67,0	76,6	78,3	71,3
i-Tec HS-741	77,4	77,0	76,2	69,0	77,3	76,7	75,6
Sennheiser PC 111	78,2	59,4	68,5	75,0	71,1	65,6	69,6
i-Tec HS-741	80,3	61,5	76,7	77,0	77,8	66,4	73,3
Sennheiser PC 131	67,1	53,2	67,8	77,1	75,2	69,0	68,3
i-Tec HS-741	71,2	70,5	71,8	81,9	82,4	65,5	73,9
Sennheiser PC 161	78,9	72,0	81,9	77,1	78,5	82,1	78,4
i-Tec HS-741	83,7	67,1	76,1	73,0	84,1	78,1	77,0

Testování s modely **trifonů**:

	Míra úspěšnosti Acc [%]						
	M - JB	M - JT	M - PZ	Z - MK	Z - PR	Z - VM	Průměr
Genius HS-02N	81,25	81,88	85,52	90,10	84,43	86,96	85,0
i-Tec HS-741	86,25	84,38	83,45	89,11	89,34	86,96	86,3
KOSS CS/100	91,24	77,58	82,78	84,00	88,31	86,67	84,9
i-Tec HS-741	87,59	83,64	81,46	83,00	89,61	90,00	85,9
Sennheiser PC 111	84,51	80,42	92,47	84,00	86,67	77,60	84,5
i-Tec HS-741	89,44	81,82	90,41	90,00	90,37	79,20	86,9
Sennheiser PC 131	79,45	80,58	81,21	83,81	88,80	80,53	82,2
i-Tec HS-741	86,99	83,45	84,56	85,71	88,80	82,30	85,3
Sennheiser PC 161	88,44	84,62	92,75	94,07	86,92	82,93	88,3
i-Tec HS-741	86,39	83,92	92,75	91,53	89,72	91,06	89,0

Příloha C – tabulka výsledků rozpoznávání pro různé polohy mikrofonů

umístění sluchátek s mikrofonem												
na hlavě						kolem krku					
nahrávka	Corr [%]	Acc [%]	nahrávka	Corr [%]	Acc [%]	nahrávka	Corr [%]	Acc [%]	nahrávka	Corr [%]	Acc [%]
1	88,9	83,3	51	100,0	100,0	101	93,3	93,3	151	100,0	100,0
2	94,7	94,7	52	100,0	100,0	102	75,0	75,0	152	100,0	100,0
3	69,2	69,2	53	100,0	100,0	103	100,0	100,0	153	100,0	100,0
4	100,0	100,0	54	92,9	92,9	104	93,8	93,8	154	93,8	93,8
5	100,0	100,0	55	61,5	30,8	105	100,0	100,0	155	85,7	85,7
6	94,1	94,1	56	69,2	69,2	106	87,5	87,5	156	80,0	80,0
7	100,0	100,0	57	100,0	100,0	107	100,0	100,0	157	60,0	60,0
8	93,3	86,7	58	61,5	61,5	108	94,7	94,7	158	94,4	94,4
9	40,0	40,0	59	100,0	100,0	109	95,0	90,0	159	100,0	90,9
10	92,3	92,3	60	100,0	100,0	110	94,1	94,1	160	83,3	83,3
11	100,0	100,0	61	93,8	93,8	111	100,0	100,0	161	89,5	89,5
12	100,0	100,0	62	100,0	100,0	112	100,0	100,0	162	100,0	100,0
13	100,0	100,0	63	100,0	100,0	113	83,3	83,3	163	100,0	100,0
14	100,0	100,0	64	100,0	100,0	114	92,9	92,9	164	91,7	91,7
15	88,9	88,9	65	100,0	100,0	115	100,0	100,0	165	87,5	87,5
16	100,0	100,0	66	72,7	72,7	116	100,0	100,0	166	100,0	100,0
17	95,0	95,0	67	80,0	80,0	117	100,0	100,0	167	88,9	88,9
18	93,8	93,8	68	96,0	96,0	118	78,9	78,9	168	100,0	100,0
19	87,5	87,5	69	100,0	100,0	119	100,0	100,0	169	100,0	100,0
20	94,1	94,1	70	84,6	84,6	120	81,8	72,7	170	93,3	93,3
21	100,0	100,0	71	100,0	100,0	121	87,5	87,5	171	80,0	66,7
22	85,7	85,7	72	100,0	100,0	122	100,0	100,0	172	100,0	100,0
23	81,3	81,3	73	100,0	100,0	123	100,0	100,0	173	100,0	100,0
24	91,7	91,7	74	100,0	100,0	124	100,0	100,0	174	100,0	100,0
25	100,0	100,0	75	100,0	100,0	125	91,7	75,0	175	87,5	87,5
26	100,0	100,0	76	86,7	86,7	126	100,0	100,0	176	100,0	100,0
27	80,0	80,0	77	90,9	90,9	127	93,3	93,3	177	95,5	95,5
28	90,0	90,0	78	71,4	71,4	128	100,0	100,0	178	75,0	75,0
29	94,1	94,1	79	100,0	100,0	129	77,8	77,8	179	100,0	100,0
30	100,0	100,0	80	100,0	100,0	130	100,0	100,0	180	75,0	75,0
31	87,5	81,3	81	85,7	85,7	131	37,5	37,5	181	88,9	88,9
32	100,0	100,0	82	100,0	100,0	132	92,3	92,3	182	93,3	86,7
33	90,0	90,0	83	94,4	94,4	133	100,0	100,0	183	88,5	84,6
34	94,1	94,1	84	100,0	100,0	134	91,7	91,7	184	91,7	91,7
35	100,0	100,0	85	100,0	83,3	135	100,0	100,0	185	93,3	93,3
36	100,0	100,0	86	93,8	87,5	136	77,8	77,8	186	85,7	85,7
37	84,2	84,2	87	100,0	100,0	137	95,0	95,0	187	100,0	100,0
38	90,9	90,9	88	88,2	88,2	138	100,0	100,0	188	88,9	88,9
39	85,7	85,7	89	100,0	100,0	139	100,0	100,0	189	85,7	78,6
40	78,6	78,6	90	90,9	81,8	140	80,0	60,0	190	100,0	100,0
41	80,0	80,0	91	94,1	94,1	141	100,0	100,0	191	100,0	100,0
42	92,9	92,9	92	100,0	100,0	142	100,0	100,0	192	100,0	100,0
43	100,0	100,0	93	80,0	60,0	143	93,8	87,5	193	90,5	90,5
44	87,5	87,5	94	100,0	100,0	144	100,0	100,0	194	72,7	72,7
45	100,0	100,0	95	92,3	92,3	145	85,7	78,6	195	50,0	50,0
46	85,7	57,1	96	100,0	100,0	146	93,3	93,3	196	100,0	100,0
47	82,4	82,4	97	95,0	85,0	147	90,0	90,0	197	100,0	100,0
48	100,0	100,0	98	100,0	100,0	148	87,5	87,5	198	100,0	100,0
49	83,3	83,3	99	81,8	81,8	149	72,7	72,7	199	90,0	90,0
50	100,0	100,0	100	100,0	100,0	150	100,0	90,9			
			průměr:	91,8	90,6				průměr:	92,3	91,1

Příloha D – tabulky výsledků rozpoznávání s monofonovými modely adaptovanými intuitivní adaptační metodou – testování proběhlo v systému vytvořeném v prostředí HTK

0xADAPT	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	63,3	62,0	52,1	52,1	57,3	57,3	57,9	57,9	47,5	42,4	62,7	60,0	56,9	55,5
i-Tec HS-741	59,5	57,0	55,2	54,2	54,9	53,7	65,8	65,8	55,9	55,9	61,3	61,3	58,0	57,1
KOSS CS/100	74,3	64,9	33,0	26,0	47,7	42,0	40,5	29,7	65,9	61,5	66,7	61,4	54,4	47,7
i-Tec HS-741	73,0	68,9	55,0	53,0	50,0	50,0	56,8	56,8	62,6	60,4	66,7	64,9	60,2	58,4
Sennheiser PC 111	51,8	49,4	40,5	39,2	42,2	41,0	54,1	54,1	58,3	56,9	37,1	33,9	46,9	45,2
i-Tec HS-741	60,0	55,3	62,0	58,2	45,8	44,6	54,1	54,1	56,9	55,6	64,5	64,5	57,2	55,0
Sennheiser PC 131	66,3	53,0	47,4	43,4	51,2	46,5	64,3	64,3	71,0	69,4	66,0	62,0	59,9	54,6
i-Tec HS-741	67,5	63,9	57,9	57,9	57,0	57,0	52,4	52,4	61,3	61,3	50,0	48,0	58,6	57,6
Sennheiser PC 161	67,9	64,3	63,8	61,3	68,9	63,5	69,1	67,3	54,5	52,3	66,7	61,7	65,7	62,2
i-Tec HS-741	70,2	66,7	63,8	56,3	68,9	68,9	49,1	49,1	54,5	54,5	48,3	46,7	60,7	58,2

1xADAPT	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	67,1	67,1	55,2	54,2	52,4	52,4	52,6	52,6	40,7	39,0	66,7	64,0	56,6	55,7
i-Tec HS-741	58,2	54,4	56,3	56,3	62,2	62,2	65,8	65,8	61,0	57,6	64,0	64,0	60,6	59,4
KOSS CS/100	81,1	70,3	33,0	29,0	46,6	43,2	45,9	43,2	63,7	58,2	64,9	64,9	55,0	50,3
i-Tec HS-741	70,3	68,9	46,0	44,0	58,0	58,0	56,8	56,8	62,6	60,4	64,9	64,9	59,1	57,9
Sennheiser PC 111	57,6	57,6	39,2	36,7	47,0	44,6	51,4	51,4	55,6	54,2	45,2	45,2	49,3	48,1
i-Tec HS-741	61,2	55,3	59,5	58,2	45,8	45,8	48,6	48,6	63,9	63,9	69,4	69,4	58,4	56,9
Sennheiser PC 131	65,1	54,2	47,4	43,4	60,5	59,3	59,5	59,5	66,1	64,5	68,0	64,0	60,7	56,6
i-Tec HS-741	66,3	61,4	56,6	56,6	61,6	60,5	59,5	59,5	62,9	62,9	52,0	50,0	60,4	58,9
Sennheiser PC 161	72,6	71,4	60,0	58,8	70,3	70,3	69,1	69,1	61,4	61,4	60,0	55,0	66,0	64,7
i-Tec HS-741	69,0	67,9	61,3	56,3	70,3	70,3	56,4	56,4	59,1	59,1	55,0	53,3	62,7	61,2

2xADAPT	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	67,1	65,8	54,2	54,2	53,7	53,7	60,5	60,5	42,4	40,7	68,0	65,3	57,8	56,9
i-Tec HS-741	60,8	58,2	56,3	56,3	62,2	62,2	65,8	65,8	64,4	61,0	66,7	66,7	62,0	61,1
KOSS CS/100	79,7	70,3	37,0	32,0	48,9	46,6	45,9	43,2	63,7	58,2	66,7	66,7	56,4	51,9
i-Tec HS-741	71,6	70,3	49,0	48,0	59,1	59,1	56,8	56,8	63,7	61,5	64,9	64,9	60,4	59,5
Sennheiser PC 111	60,0	57,6	38,0	35,4	48,2	45,8	45,9	45,9	55,6	54,2	53,2	53,2	50,5	48,8
i-Tec HS-741	63,5	55,3	58,2	54,4	54,2	54,2	48,6	48,6	62,5	62,5	69,4	69,4	60,0	57,7
Sennheiser PC 131	65,1	57,8	50,0	46,1	62,8	61,6	64,3	64,3	71,0	69,4	68,0	66,0	62,9	59,9
i-Tec HS-741	67,5	62,7	56,6	56,6	59,3	59,3	59,5	59,5	62,9	62,9	52,0	50,0	60,2	58,9
Sennheiser PC 161	71,4	70,2	62,5	60,0	73,0	71,6	67,3	67,3	61,4	61,4	60,0	56,7	66,5	65,0
i-Tec HS-741	67,9	64,3	61,3	56,3	67,6	67,6	58,2	58,2	59,1	59,1	51,7	51,7	61,7	59,9

4xADAPT	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	68,4	67,1	54,2	54,2	56,1	56,1	63,2	60,5	47,5	45,8	68,0	65,3	59,4	58,3
i-Tec HS-741	59,5	55,7	56,3	56,3	59,8	59,8	63,2	63,2	66,1	62,7	66,7	66,7	61,3	60,1
KOSS CS/100	81,1	71,6	41,0	33,0	52,3	48,9	54,1	51,4	65,9	60,4	71,9	71,9	60,0	54,6
i-Tec HS-741	74,3	70,3	51,0	49,0	52,3	52,3	56,8	56,8	65,9	63,7	61,4	61,4	60,0	58,4
Sennheiser PC 111	57,6	55,3	41,8	38,0	53,0	50,6	45,9	45,9	56,9	56,9	53,2	51,6	51,9	50,0
i-Tec HS-741	64,7	56,5	64,6	62,0	53,0	53,0	54,1	54,1	63,9	62,5	71,0	71,0	62,2	59,8
Sennheiser PC 131	68,7	63,9	51,3	46,1	61,6	60,5	61,9	61,9	74,2	74,2	68,0	66,0	63,9	61,4
i-Tec HS-741	67,5	62,7	60,5	60,5	60,5	60,5	59,5	59,5	62,9	62,9	52,0	50,0	61,2	59,9
Sennheiser PC 161	71,4	66,7	60,0	57,5	71,6	70,3	67,3	67,3	61,4	61,4	56,7	55,0	65,2	63,2
i-Tec HS-741	67,9	64,3	65,0	60,0	71,6	70,3	56,4	56,4	63,6	63,6	55,0	55,0	64,0	62,0

8xADAPT	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	68,4	65,8	62,5	62,5	62,2	62,2	60,5	52,6	52,5	50,8	70,7	69,3	63,4	61,8
i-Tec HS-741	54,4	50,6	56,3	56,3	63,4	63,4	60,5	60,5	62,7	59,3	69,3	69,3	60,8	59,7
KOSS CS/100	81,1	73,0	48,0	40,0	53,4	50,0	54,1	51,4	69,2	64,8	71,9	70,2	62,4	57,3
i-Tec HS-741	73,0	68,9	53,0	50,0	58,0	56,8	54,1	54,1	65,9	63,7	63,2	63,2	61,3	59,3
Sennheiser PC 111	60,0	55,3	46,8	44,3	54,2	51,8	45,9	45,9	58,3	58,3	58,1	56,5	54,5	52,4
i-Tec HS-741	61,2	54,1	62,0	58,2	57,8	56,6	56,8	56,8	66,7	65,3	67,7	67,7	62,2	59,6
Sennheiser PC 131	65,1	59,0	57,9	55,3	57,0	55,8	64,3	64,3	66,1	66,1	70,0	68,0	62,7	60,4
i-Tec HS-741	66,3	61,4	60,5	57,9	61,6	61,6	59,5	59,5	67,7	67,7	52,0	50,0	61,9	60,2
Sennheiser PC 161	72,6	70,2	63,8	61,3	75,7	74,3	69,1	69,1	59,1	59,1	60,0	58,3	67,5	66,0
i-Tec HS-741	66,7	61,9	73,8	68,8	71,6	66,2	60,0	60,0	61,4	61,4	56,7	56,7	66,0	63,0

Příloha E – tabulky výsledků rozpoznávání adaptovanými modely v systému rozpoznávání vytvořeném v prostředí HTK

NO adapt	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	74,7	72,2	60,4	59,4	73,2	73,2	65,8	65,8	64,4	62,7	68,0	68,0	67,8	66,9
i-Tec HS-741	74,7	72,2	64,6	62,5	72,0	72,0	73,7	73,7	71,2	69,5	70,7	70,7	70,6	69,5
KOSS CS/100	83,8	82,4	53,0	49,0	67,0	67,0	75,7	75,7	79,1	79,1	75,4	73,7	70,9	69,6
i-Tec HS-741	86,5	85,1	70,0	66,0	67,0	65,9	73,0	70,3	74,7	73,6	75,4	75,4	74,0	72,3
Sennheiser PC111	74,1	72,9	58,2	57,0	67,5	67,5	59,5	59,5	66,7	66,7	71,0	69,4	66,7	66,0
i-Tec HS-741	80,0	78,8	67,1	64,6	66,3	66,3	62,2	62,2	68,1	68,1	74,2	74,2	70,3	69,6
Sennheiser PC131	75,9	73,5	60,5	59,2	64,0	61,6	69,0	69,0	80,6	80,6	72,0	72,0	69,9	68,7
i-Tec HS-741	75,9	72,3	69,7	69,7	64,0	62,8	66,7	66,7	80,6	80,6	76,0	76,0	71,9	70,9
Sennheiser PC161	77,4	76,2	70,0	70,0	78,4	75,7	74,5	72,7	68,2	68,2	71,7	70,0	73,8	72,5
i-Tec HS-741	82,1	81,0	73,8	71,3	78,4	75,7	63,6	63,6	65,9	65,9	66,7	66,7	73,0	71,8

MAP adapt	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	74,7	70,9	62,5	60,4	74,4	74,4	65,8	65,8	66,1	64,4	68,0	68,0	68,8	67,4
i-Tec HS-741	77,2	74,7	68,8	66,7	70,7	70,7	76,3	76,3	71,2	69,5	70,7	70,7	72,0	70,9
KOSS CS/100	83,8	82,4	59,0	55,0	69,3	69,3	70,3	70,3	79,1	79,1	75,4	75,4	72,3	71,1
i-Tec HS-741	85,1	82,4	71,0	67,0	67,0	65,9	73,0	70,3	78,0	76,9	75,4	75,4	74,7	72,7
Sennheiser PC111	74,1	72,9	59,5	58,2	66,3	66,3	59,5	59,5	66,7	66,7	71,0	69,4	66,7	66,0
i-Tec HS-741	80,0	78,8	67,1	64,6	63,9	63,9	64,9	64,9	69,4	69,4	75,8	75,8	70,6	69,9
Sennheiser PC131	75,9	73,5	60,5	59,2	66,3	64,0	64,3	64,3	80,6	80,6	70,0	70,0	69,7	68,4
i-Tec HS-741	78,3	74,7	71,1	71,1	65,1	65,1	66,7	66,7	80,6	80,6	78,0	78,0	73,2	72,4
Sennheiser PC161	76,2	75,0	70,0	70,0	78,4	75,7	76,4	76,4	65,9	65,9	66,7	66,7	72,8	72,0
i-Tec HS-741	82,1	81,0	73,8	72,5	77,0	74,3	67,3	67,3	65,9	65,9	68,3	68,3	73,6	72,5

MLLR adapt	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	77,2	75,9	69,8	67,7	68,3	68,3	65,8	65,8	71,2	69,5	70,7	70,7	70,9	69,9
i-Tec HS-741	74,7	70,9	68,8	66,7	69,5	69,5	81,6	81,6	74,6	71,2	70,7	70,7	72,3	70,6
KOSS CS/100	85,1	82,4	55,0	48,0	70,5	69,3	78,4	78,4	81,3	81,3	75,4	71,9	72,9	70,2
i-Tec HS-741	89,2	87,8	72,0	68,0	75,0	73,9	75,7	73,0	78,0	78,0	75,4	75,4	77,4	75,8
Sennheiser PC111	77,6	76,5	62,0	60,8	63,9	63,9	59,5	59,5	68,1	68,1	74,2	72,6	68,2	67,5
i-Tec HS-741	84,7	83,5	68,4	65,8	69,9	69,9	67,6	67,6	72,2	72,2	71,0	71,0	73,0	72,2
Sennheiser PC131	80,7	78,3	65,8	64,5	64,0	62,8	69,0	69,0	75,8	75,8	74,0	74,0	71,4	70,4
i-Tec HS-741	77,1	72,3	76,3	75,0	67,4	67,4	69,0	69,0	79,0	79,0	76,0	76,0	74,2	72,9
Sennheiser PC161	76,2	75,0	76,3	76,3	78,4	75,7	72,7	72,7	65,9	65,9	66,7	65,0	73,6	72,5
i-Tec HS-741	86,9	85,7	73,8	71,3	81,1	79,7	67,3	67,3	65,9	65,9	70,0	70,0	75,6	74,6

MAP + MLLR adapt	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	77,2	75,9	69,8	67,7	68,3	68,3	65,8	65,8	67,8	66,1	70,7	70,7	70,4	69,5
i-Tec HS-741	77,2	74,7	69,8	67,7	69,5	69,5	81,6	81,6	76,3	72,9	69,3	69,3	73,0	71,6
KOSS CS/100	85,1	82,4	55,0	48,0	73,9	72,7	78,4	78,4	80,2	80,2	75,4	71,9	73,4	70,7
i-Tec HS-741	89,2	87,8	74,0	70,0	73,9	72,7	75,7	73,0	78,0	76,9	75,4	75,4	77,6	75,8
Sennheiser PC111	78,8	77,6	62,0	60,8	61,4	61,4	59,5	59,5	68,1	68,1	74,2	72,6	67,9	67,2
i-Tec HS-741	85,9	84,7	68,4	67,1	69,9	69,9	67,6	67,6	72,2	72,2	72,6	72,6	73,4	73,0
Sennheiser PC131	80,7	78,3	64,5	63,2	65,1	64,0	71,4	71,4	80,6	80,6	74,0	74,0	72,4	71,4
i-Tec HS-741	78,3	72,3	76,3	75,0	67,4	67,4	69,0	69,0	79,0	79,0	80,0	80,0	74,9	73,4
Sennheiser PC161	79,8	78,6	73,8	73,8	78,4	75,7	72,7	72,7	65,9	65,9	66,7	65,0	73,8	72,8
i-Tec HS-741	88,1	86,9	75,0	72,5	81,1	79,7	67,3	67,3	68,2	68,2	70,0	70,0	76,3	75,3

Příloha F – tabulky výsledků rozpoznávání adaptovanými modely v systému rozpoznávání poskytnutém Laboratoří počítačového zpracování řeči

NO adapt	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	83,8	83,8	80,2	79,2	77,8	76,5	89,2	89,2	77,6	75,9	86,5	86,5	81,9	81,2
i-Tec HS-741	90,0	87,5	77,1	75,0	81,5	80,2	91,9	91,9	91,4	91,4	87,8	87,8	85,4	84,3
KOSS CS/100	93,2	91,8	78,2	73,3	92,0	89,7	80,6	80,6	90,0	87,8	94,6	94,6	88,0	85,8
i-Tec HS-741	89,0	86,3	87,1	84,2	83,9	82,8	80,6	80,6	91,1	88,9	94,6	94,6	88,0	86,2
Sennheiser PC111	94,0	90,5	81,0	77,2	91,5	87,8	80,6	80,6	81,7	81,7	67,2	62,3	83,8	80,9
i-Tec HS-741	94,0	94,0	81,0	77,2	92,7	89,0	80,6	80,6	87,3	87,3	73,8	70,5	86,0	84,0
Sennheiser PC131	86,6	80,5	78,7	78,7	85,9	83,5	85,4	85,4	85,2	80,3	85,7	83,7	84,5	81,7
i-Tec HS-741	87,8	82,9	85,3	85,3	85,9	84,7	87,8	85,4	88,5	85,2	85,7	83,7	86,8	84,5
Sennheiser PC161	88,0	86,7	82,3	82,3	91,9	91,9	85,2	85,2	81,4	76,7	91,5	91,5	87,0	86,2
i-Tec HS-741	91,6	90,4	81,0	79,7	94,6	94,6	90,7	88,9	86,0	86,0	86,4	86,4	88,5	87,8

MAP adapt	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	82,5	82,5	78,1	77,1	84,0	82,7	86,5	86,5	79,3	77,6	86,5	86,5	82,4	81,7
i-Tec HS-741	88,8	85,0	77,1	75,0	81,5	80,2	91,9	91,9	89,7	89,7	87,8	87,8	85,0	83,6
KOSS CS/100	94,5	94,5	77,2	68,3	89,7	88,5	80,6	80,6	87,8	86,7	94,6	94,6	87,1	84,7
i-Tec HS-741	89,0	84,9	86,1	82,2	82,8	80,5	80,6	80,6	91,1	88,9	94,6	94,6	87,6	85,1
Sennheiser PC111	94,0	91,7	81,0	77,2	91,5	87,8	80,6	80,6	85,9	85,9	65,6	60,7	84,3	81,6
i-Tec HS-741	95,2	95,2	83,5	79,7	92,7	90,2	77,8	77,8	87,3	87,3	73,8	70,5	86,4	84,7
Sennheiser PC131	84,1	79,3	78,7	77,3	85,9	83,5	85,4	85,4	86,9	83,6	85,7	83,7	84,2	81,7
i-Tec HS-741	85,4	81,7	85,3	85,3	84,7	83,5	87,8	85,4	88,5	85,2	85,7	83,7	86,0	84,0
Sennheiser PC161	90,4	89,2	82,3	82,3	91,9	91,9	85,2	85,2	81,4	76,7	91,5	91,5	87,5	86,7
i-Tec HS-741	91,6	90,4	81,0	81,0	94,6	94,6	90,7	88,9	86,0	86,0	86,4	86,4	88,5	88,0

MLLR adapt	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	93,8	91,3	84,4	82,3	82,7	82,7	81,1	81,1	81,0	79,3	86,5	86,5	85,4	84,3
i-Tec HS-741	92,5	88,8	81,3	80,2	81,5	77,8	91,9	91,9	87,9	86,2	87,8	87,8	86,4	84,5
KOSS CS/100	95,9	93,2	83,2	77,2	92,0	90,8	75,0	72,2	90,0	88,9	94,6	94,6	89,2	86,7
i-Tec HS-741	93,2	89,0	88,1	86,1	89,7	89,7	83,3	83,3	94,4	92,2	96,4	96,4	91,2	89,6
Sennheiser PC111	94,0	92,9	84,8	81,0	95,1	93,9	80,6	80,6	88,7	87,3	73,8	68,9	87,4	85,2
i-Tec HS-741	95,2	95,2	86,1	78,5	92,7	90,2	83,3	83,3	88,7	87,3	75,4	73,8	87,9	85,5
Sennheiser PC131	90,2	82,9	85,3	85,3	81,2	80,0	87,8	87,8	88,5	83,6	85,7	85,7	86,3	83,7
i-Tec HS-741	89,0	85,4	82,7	81,3	87,1	85,9	80,5	80,5	90,2	85,2	85,7	83,7	86,3	84,0
Sennheiser PC161	86,7	84,3	77,2	72,2	93,2	93,2	88,9	88,9	83,7	81,4	91,5	91,5	86,7	84,9
i-Tec HS-741	90,4	88,0	87,3	86,1	95,9	95,9	90,7	88,9	86,0	86,0	86,4	86,4	89,8	88,8

MAP + MLLR adapt	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	91,3	88,8	81,3	80,2	80,2	80,2	81,1	81,1	79,3	77,6	86,5	86,5	83,6	82,6
i-Tec HS-741	92,5	88,8	81,3	80,2	80,2	76,5	91,9	91,9	87,9	86,2	87,8	87,8	86,2	84,3
KOSS CS/100	95,9	95,9	84,2	79,2	93,1	92,0	75,0	72,2	92,2	91,1	94,6	94,6	90,1	88,3
i-Tec HS-741	93,2	89,0	90,1	88,1	89,7	89,7	83,3	83,3	96,7	94,4	96,4	96,4	92,1	90,5
Sennheiser PC111	94,0	92,9	83,5	79,7	96,3	95,1	80,6	80,6	88,7	87,3	75,4	72,1	87,7	85,7
i-Tec HS-741	94,0	94,0	87,3	79,7	93,9	90,2	83,3	83,3	88,7	87,3	75,4	73,8	88,1	85,5
Sennheiser PC131	92,7	87,8	85,3	85,3	87,1	84,7	87,8	87,8	88,5	83,6	85,7	85,7	88,0	85,8
i-Tec HS-741	86,6	82,9	82,7	81,3	87,1	85,9	80,5	80,5	90,2	85,2	85,7	83,7	85,8	83,5
Sennheiser PC161	88,0	85,5	78,5	73,4	93,2	93,2	88,9	88,9	83,7	81,4	91,5	91,5	87,2	85,5
i-Tec HS-741	91,6	89,2	89,9	89,9	95,9	95,9	90,7	88,9	86,0	86,0	86,4	86,4	90,6	89,8

CMLLR adapt	mluvčí JB [%]		mluvčí JT [%]		mluvčí PZ [%]		mluvčí MK [%]		mluvčí PR [%]		mluvčí VM [%		Průměr [%]	
	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc	Corr	Acc
Genius HS-02N	81,3	76,3	85,4	84,4	84,0	84,0	91,9	91,9	79,3	77,6	86,5	86,5	84,3	82,9
i-Tec HS-741	86,3	81,3	81,3	81,3	85,2	85,2	97,3	94,6	93,1	91,4	86,5	86,5	86,9	85,4
KOSS CS/100	95,9	95,9	83,2	81,2	92,0	90,8	83,3	80,6	90,0	87,8	94,6	94,6	89,8	88,5
i-Tec HS-741	93,2	89,0	86,1	84,2	87,4	86,2	83,3	83,3	93,3	90,0	92,9	91,1	89,6	87,4
Sennheiser PC111	97,6	95,2	87,3	86,1	95,1	93,9	86,1	86,1	90,1	87,3	80,3	78,7	90,3	88,6
i-Tec HS-741	98,8	98,8	89,9	87,3	93,9	91,5	88,9	88,9	95,8	93,0	73,8	73,8	91,0	89,6
Sennheiser PC131	89,0	82,9	86,7	85,3	83,5	81,2	85,4	85,4	91,8	88,5	87,8	81,6	87,3	84,0
i-Tec HS-741	89,0	82,9	81,3	80,0	85,9	83,5	90,2	90,2	90,2	85,2	85,7	81,6	86,8	83,5
Sennheiser PC161	88,0	85,5	81,0	81,0	95,9	94,6	88,9	88,9	83,7	81,4	86,4	86,4	87,5	86,5
i-Tec HS-741	90,4	89,2	87,3	84,8	95,9	95,9	92,6	90,7	88,4	88,4	86,4	86,4	90,3	89,3